Journal of Research in Education

Fall 2003
An International Journal Published
by the Eastern Educational Research Association

CONTENTS

Promoting Teacher Professional Judgment

Alan J. Reiman
and
Lisa Johnson
North Carolina State University

The field of teacher education has witnessed renewed attention to teacher "dispositions." Yet dispositions are poorly defined and little research evidence exists to guide teacher education and professional development programs. This paper addresses these challenges. First, the authors review the literature for insights into a working definition of teachers' dispositions. Second, they describe theory-driven interventions and related evidence from 12 studies designed to promote changes in dispositions of prospective and practicing teachers. Effect sizes are reported across three dispositional domains – moral/ethical, ego, and conceptual judgment. Effect sizes are significant for the interventions and range from +.75 (N=10, moral/ethical judgment), to+.59 (N=3, ego judgment), and +.50 (N=9, conceptual judgment). Implications are drawn for teacher education and professional development programs.

Recently, we have seen renewed attention in teacher "dispositions," as expressed by the National Council for the Accreditation of Teacher Education (NCATE) 2000 standards, the National Board for Professional Teaching Standards (1998), and the Interstate New Teacher Assessment and Support Consortium (INTASC) (1992). NCATE views effective teaching as the intersection of teachers' knowledge, skills, and dispositions. Dispositions are described as the bridge between knowledge and skills. Yet dispositions are poorly defined in the literature and there is little research evidence regarding intentional programs that can guide teacher education and teacher professional development programs interested in fostering teacher dispositions. This paper addresses these challenges. First, we review the literature and then provide a working definition. Then we introduce seven guiding theoretical principles for design of Deliberative Psychological and Professional Development (DPPE) interventions and related evidence from twelve studies designed to promote positive changes in dispositions (professional judgments) of prospective and practicing teachers. Effect sizes are reported across three dispositional domains – moral/ethical judgment (action choices in situations where moral claims conflict), ego reasoning (interpersonal relations and character), and conceptual/level (sensitivity to interpersonal relationships and problem solving). Relationships between these domains and professional action are described. Finally, implications for research and program development in teacher education are suggested.

Statement of the Problem

The use of disposition as a measure of teacher effectiveness presents three challenges: a need for clear definition; a need for supporting evidence of its importance; and a need to demonstrate that such dispositions can be positively changed in professional education programs. These issues are now addressed.

Selected Review of Literature

Teacher dispositions have reemerged in teacher education, teacher professional development, and standards-based reforms (Andrews & Schwab, 1995). However, dispositions have a variety of meanings in teacher education literature. As well, there is little research evidence regarding intentional programs interested in fostering teacher dispositions. In this literature review, we explore the related problems of lack of definition of disposition and lack of evidence regarding how to foster positive changes in novice teachers' dispositions. We propose that professional judgment and action should be central features of teacher disposition.

Need for Definition

Inquiry into teacher dispositions can be traced back at least as far as John Dewey (1904). Dewey claimed that teachers face three central challenges: (1) mastery of subject

matter and the principles of education (e.g., responsiveness to learners); (2) growth of competence in instructional and classroom management skills; and (3) development of dispositions toward inquiry, reflection, ethical judgment, and an orientation to direct attention to the underlying intellectual, affective, and motivational processes of students (Dewey, 1904). More recently, such works as Perceiving, Behaving, and Becoming (Association of Supervision and Curriculum Development, 1962) offer insights into the importance of dispositions. Contributors such as Kelly, Rogers, Maslow, and Combs offer a social-psychological view of human potential. Their basic assumption being that a truly responsible teacher has a positive and accurate view of self, is creative and open to new learning and complex new experiences, and behaves with conviction, principles, and integrity in professional contexts.

Yet, the current discussion of dispositions in teaching and teacher education appears to be ahistorical and definitions of disposition are illusive. Typically, there is a broad definition that includes a host of personal qualities or characteristics that teachers should or do possess. Work by Katz and Raths (1985) and Raths (2000) have been important in clarifying the definition of disposition by contrasting it with potentially confusing terms. For example, they contrast beliefs and dispositions by citing work from Kennedy (1997). He notes that teacher candidates bring a set of beliefs into teacher education programs and these beliefs become a force against change in the candidates' practice. One belief that teacher candidates bring to their professional schooling is "that they already have what it takes to be a good teacher, and that therefore they have little to learn from formal study of teaching" (Kennedy, p.14). Raths (2000) points out that if teachers' beliefs can interfere with their own learning or with their instruction and students' learning, then teacher educators should change these beliefs. Yet "we are not sure which are 'better' beliefs, and if we knew, we do not have a way of changing them" (Raths, 2000, p.389).

Dispositions as Professional Judgments and Actions

We propose that professional interpretation, judgment and action should be the one core characterizations of teacher disposition. Shulman (1998) highlights the central role of professional judgment. Judgment is an act of deciding. As Shulman notes, professional judgment is a disposition toward deciding based on careful consideration of theoretical knowledge and skills with the needs of learners. As such, it creates bridges between the universal terms of theory and the "gritty particularities of situated practice" (Shulman, 1998, p. 519). Such judgment always incorporates the moral, the reflective, and the social (Mentkowski & Associates, 2000; Oser, Dick, & Patry, 1992). "The starting point for professional preparation is the premise that the aims of professionalism involve social purposes and responsibilities that are both technically and morally grounded" (Shulman, 1998, p. 516).

In light of our review of literature and standards of professional educational organizations (Interstate New Teacher Assessment and Support Consortium Standards 1992; National Board for Professional Teaching Standards, 1998), we propose the following working definition of professional disposition:

> Professional disposition is defined as an attributed characteristic of a teacher that represents a trend of a teacher's interpretation, judgments and actions in ill-structured and progressively more complex laboratory and professional contexts. Further, it is assumed that these dispositions (i.e., trends in teacher interpretations, judgments, and actions), develop over time when there is high quality professional education.

We now turn to the related problem of lack of evidence to guide teacher education programs hoping to foster positive changes in dispositions

Need for Evidence

There is a need for resarch evidence regarding teacher education programs that can foster positive changes in teacher dispositions. There is little solid evidence suggesting that teacher education programs are promoting changes. For example, Tatto (1996) reports that most teacher education programs, as they are now structured, do not change the belief systems of prospective teachers. Values clarification programs are also ineffective in fostering more ethical and caring judgments (Leming, 1981). Recent efforts to create professional development schools as envisioned by the Holmes Group (1995), though laudable, have not generated a strand of evidence on the disposition question. Taylor and Wasicsko (2000) have begun to design systems for assessing perceptual dispositions but this work is in its early stages. Raths (2000) notes that efforts to promote candidate's developmental growth to higher-level stages, as reflected in work such as Women's Ways of Knowing (Belenky, Clinchy, Goldberger, & Tarule, 1986), are largely idiosyncratic. Such studies provide little evidence of systemic interventions that could be applied to teacher education programs interested in fostering positive changes in teacher candidate's dispositions.

Judgment Research in Education

Research on educators' judgments in the conceptual/reflective domain offers promising insights. For example, Miller's (1981) review of over 200 studies of interactions between teachers' conceptual/reflective judgments and performance reported positive relationships. Teachers who consistently utilized more complex conceptual/reflective judgments had an ability to more effectively adapt to student needs, to take students' perspectives during classroom work, and to think on their feet and find alternative solutions. Reiman and Watson (1999) found that mentor teachers at

complex conceptual/reflective judgment levels understood the broader implications of their assistance to new teachers and were able to foster more engaging and democratic interactions with their teacher colleagues.

Research on educators' judgments in the ethical judgment domain suggests similar positive relationships to teacher interactions with students. Over 20 studies have been reviewed by Chang (1994). We summarize two studies that are representative of this larger group of studies. Johnston and Lubromov (1987) utilized case study methodology and found that teachers utilizing principled ethical judgments (postconventional ethical reasoning) encouraged students to participate in rule making, and the teachers were more willing to help students understand the reasons for rules. Oser (1992) found that teachers at principled ethical judgment levels were more apt to encourage students to take multiple perspectives. Another finding from the experimental research finds that teachers at postconventional ethical judgment levels foster "discourse two" dialogue with students. Discourse two is described as more inductive and learner-centered instruction.

In addition to the need for clear definitions of disposition and the need for supporting evidence of its importance, there is a need to demonstrate that such dispositions can be positively changed in professional education programs. The question of how to foster positive changes in teacher judgment domains is now described in the method section.

Method

We summarize data from twelve psychological and professional development interventions (DPPE) that span the teacher professional career spectrum from – pre-service teacher education through induction (first three years of teaching), and ongoing professional development. The studies were designed to promote positive changes in selected teaching or mentoring performances. As well, the studies were designed to promote changes in moral/ethical judgment, ego judgment and conceptual judgment. The use of overlapping assessments of the dependent variables – moral/ethical judgment,conceptual judgment, and ego judgment is based on the author's' conception of the teaching enterprise: teaching as a moral practice (e.g. ethical judgments and actions); teaching as ego understanding (e.g., allocentric and caring practice requiring a self-actualized person); and teaching as practice of professional (conceptual) judgments — optimal evaluation and justification of professional decisions and assumptions regarding classroom management and needs of diverse learners. The research summary includes the following assumptions:

- There is a consistent predictive relationship between more complex judgment reasoning and performance in complex "human-helping" tasks like teaching.
- The moral/ethical, conceptual epistemological/ conceptual, and ego domains of complexity may inform teacher educators and policymakers about the nature of dispositions that which are now acknowledged as central to teacher competence and expertise.
- The professional judgment levels of pre-service teachers and beginning teachers are modest, about the mid-point on the judgment indices used in this research summary.
- Educational programming should assess how professional education and ongoing professional development fosters growth across the moral/ ethical, conceptual, and ego judgment domains.

Design

The meta-analysis includes the entire population of eligible research studies that have used the DPPE model as the independent variable (Lipsey & Wilson, 1993). Following the guidelines of Cooper and Hedges (1994) and Rosenthal (1995), the investigators obtained all experimental or quasi-experimental studies published in English-language journals or as dissertations from 1972 to 2001. This search produced twelve studies. Selection criteria ensured, to the extent possible, that the studies in the final sample would reflect all studies that had utilized the DPPE model. The studies met the criteria:

- The study had to be experimental or quasi-experimental in design.
- The study had to incorporate the DPPE model.
- Dependent measures needed to include the assessment of judgment in various teacher career phases or in pre-service teacher education.
- The post-test measure/s needed to occur at least three months after the pre-test.
- The study had to be a doctoral dissertation or be published in an English-language peer reviewed journal.
- The study had to include teachers or related educational professionals (e.g. counselors).

A number of studies were excluded from this initial list because one or more of these criteria were not met.

Independent Variable

How was the DPPE operationalized as an independent variable in the studies reported? The original start on the intervention question was in secondary schools. These interventions (Mosher & Sprinthall, 1970) were guided by Mead's concept of social role-taking (1934) as a bridge or condition for adult development. More recently, Selman (1980) has described social role-taking and the growth of interpersonal understanding. Selman and his colleagues visualize a developmental nature to psychosocial competence that includes forms of shared experience and coordination of increasingly complex social perspective

taking (Selman, Watts, & Schultz, 1997).

As the cycles of research unfolded, interest grew in the framework's application to teacher education. Teacher education, like many other fields, needs an applied theory of assistance for guiding educational programming aimed at fostering positive changes in performance and disposition. Sikula (1996) noted that there are very few efforts to systematically test theory and its application in teacher education. The seven design principles have evolved as a consequence of research during the last two decades and summarized in recent research reviews(Joyce & Showers, 1995; Oja & Smulyan, 1989; Sprinthall, Reiman, & Thies-Sprinthall, 1996; Reiman, 19998). The first iteration of the principles was described by Sprinthall and Thies-Sprinthall (1983). The fundamental conditions within this theoretical/ conceptual framework for promoting positive transformations in performance and professional judgment across the moral/ ethical, conceptual, or ego domains follow:

1. *Contextualized Learning and Developmen*t. First, educators of adults must contextualize learning and instruction by accounting for the experiences of diverse learners who are taking on new and expanded professional roles. This element includes acknowledging prior knowledge/experience and developing rapport with the learners.

2. *New Role-Taking (not role playing).* Second, when teachers undertake complex new human-helping roles such as collaborative action researcher, mentor, or school-based teacher educator, the role-taking (action) precedes and shapes the intellectual consciousness that grows out of it. Guided inquiry that which includes analysis and reflection best grows out of real problems present in immediate experience of the complex new role.

3. *Guided Inquiry.* Guided inquiry includes both analysis (intensive written self-assessment using a variety of rubrics) and meta-reflection. Experiential learning can be just as arid as listening to lectures. Carefully planned activities encourage self-assessment of performance, ongoing discussion, and dialogic journaling (reflection). These assessment and reflection activities are guided by a "more capable other."

4. *Balance.* It is important that action (new role-taking) and inquiry remain in balance. Usually this means that the complex new role or helping activity is sequenced with guided inquiry each week. Too great a time lag between action and inquiry or the other way around appears to halt the growth process.

5. *Continuity.* There is a learning truism that spaced practice is vastly superior to massed practice. The complex goal of fostering both learning and development that, of course, includes ego, conceptual, and moral/ethical development, requires a continuous interplay between action and reflection. A one-or two-week workshop followed by actual helping has not caused shifts in the cognitive structures (development) of the participants. Typically, as least four to six months are needed for significant developmental changes.

6. *Support and Challenge.* The investigators essentially rediscovered Vygotsky's zone of proximal growth (1978). Support and optimal challenge (prompting the learner to accommodate to new learning) are necessary for integrated learning. This is the most complex pedagogical requirement of the DPPE approach. Persons in the complex new roles face new and complex responsibilities, and are often in the middle of a "knowledge perturbation" as they begin the new role. Teacher educators must balance support and encouragement with progressively more complex intellectual challenges that are just beyond the current preferred ways of problem solving of the novice teacher.

7. *Reflective Coaching.* Attention is given to fostering new skills. Coaching for new performances requires an instructional model, wherein the adult learner, over time, acquires "executive control" of complex new performances. Elements of reflective coaching include assessment of prior performance, overview of related theory and evidence, demonstration, opportunity for guided practice and feedback, and eventual adaptation and generalization of the performance.

Measures

Typically, the studies in this research review utilized overlapping measurement of two judgment domains. Measures of varied domains of judgment reasoning included the moral/ethical domain, the conceptual/reflective domain, and the ego (self-and social-understanding) domain.

Assessment of Conceptual Reasoning. The Paragraph Completion test was used to assess conceptual judgment. Hunt (1974) defines conceptual judgment as a personality dimension that includes cognitive complexity as well as interpersonal maturity. The most comprehensive review of content and predictive validity was conducted by Miller (1981). The review of hundreds of studies supported the idea that teachers at more complex conceptual levels need less structure in professional development programming, tend to be more responsive to student needs, and are more flexible at making changes to instruction in order that it is more facilitative of student learning.

Assessment of Moral/Ethical Reasoning. The Defining Issues Test (DIT) (Rest, 1986) assessed moral/ethical judgment. The premise of the DIT is that persons at different ages and different levels of moral complexity will interpret moral dilemmas differently. The DIT is a six dilemma objective test. Responses are scored by an outside evaluator. Without deliberative interventions, significant changes in principled

judgments do not occur. The most recent work on this test is summarized by Rest, Narvaez, Bebeau, and Thoma (1999).

Assessment of Ego (Judgments about Self and Others). The Sentence Completion Test (SCT) assesses changes in a person's judgments about intrapersonal and interpersonal relations. It is based on Loevinger's theory of ego (1976) that characterizes ego as a unity of personality, individuality, and method of facing personal problems and the problems of life. The test consists of thirty-six sentence stems. The response to each stem is independently rated as manifesting one of nine levels of ego maturity: Impulsive, Self-Protective, Conformist, Self-Aware, Conscientious, Individualistic, Autonomous, and Integrated.

Effect Size

Effect sizes are reported for the studies. Bowen (1977) has proposed the following rules of thumb for interpreting effect sizes: small = 0.10-0.39, moderate = 0.40-0.69, large = 0.70-0.99, and very large = 1.00 and above. Non-significant findings were included as well (Rosenthal, 1995). Thus, non-significant findings ($p > .05$) were included in calculations of the average effect size. Zero was substituted in the computation of averages when standard deviation data was unavailable for computing the effect size.

Settings and Limitations of the Reviewed Studies

There are a number of potential confounding variables in the reviewed studies. The first limitation involves potential interactions between the tests and the interventions. In general, exposure to the pretests does not lead to better performance on the posttests. Although this is a common concern for psychological tests, assessments of the potential of reactive effects have shown low interactions between testing and subsequent performance on the posttest. A second confounding variable is the use of volunteer samples for the greater proportion of the reviewed studies. Gage (1985) addresses the issue of external validity, noting that generalizations should only be made to comparable groups.

A more difficult limitation involves the integrity of the independent variables being assessed. The nature of the DPPE prevents the kinds of carefully controlled research that can occur in a laboratory setting. In the case of this review, the DPPE model was familiar to all investigators. However, the model has evolved since the early 1980s and now exists in a more explicit form that includes guidelines for differentiating guided inquiry. We also know that initial relationship -building in DPPEs can have dramatic effects on the outcomes thanks to research by Watson and Reiman (2001). The authors acknowledge the difficulties of field-based research to design in ways that are responsive to the learners and to the theory.

The studies that are summarized reflect teacher professional development programming in the Southeastern, Northeastern, and Midwestern sections of the United States and. employed in both urban and rural settings. The samples tended to have diverse make-up according to ethnicity, but lacked gender diversity (females were in the majority in all samples).

Data Analysis

The studies included in the meta-analysis were experimental or quasi-experimental in nature. These studies joined new social role-taking with the inquiry method (analysis and guided reflection) that was described previously in the discussion of independent variables. A summary of guided inquiry and role-taking curricular strategies is described in Reiman, Sprinthall, and Thies-Sprinthall (1999). The programs included new performances that were being learned as a consequence of being in a new role. The other conditions of support, challenge, continuity, and coaching were met.

Table 1 shows an overviews DPPE studies. In each case, the studies asked educators to assume complex new roles. As the new role was assumed, the participants engaged in a scope- and sequenced-professional development curriculum that followed the elements outlined in the DPPE framework. Guided inquiry (careful and sustained self-assessment of performance and dialogic meta-reflection) was provided during the four-month to one-year interventions. The studies show consistent positive gains across the three domains – moral/ethical, ego (self and social understanding), and conceptual (epistemology and reflectivity).

There were 10 DPPEs (11 comparison groups) that included moral/ethical judgment as a dependent variable. The average effect size across the 10 studies (11 comparisons) was +.75. This effect indicates that DPPEs led to significant positive transformations in persons' postconventional ethical judgments as measured by the Defining Issues Test. Postconventional ethical judgments have four elements: primacy of moral criteria, appeal to an ideal, sharable ideals, and full reciprocity (Rest, Narvaez, Bebeau, & Thoma, 1999). Primacy of moral criteria means that duties and rights follow from a moral purpose behind those conventions. Appeal to an ideal means that the person tends to offer constructive ideals with which to transform society. Sharable ideals means that those ideals must be sharable rather than idiosyncratic. Sharability is tested by the ability to justify an act or practice to those whose participation is expected. Full reciprocity entails uniform application of norms while recognizing that the norms themselves not be biased in favor of some at the expense of others.

Three DPPE studies utilized the Loevinger Sentence Completion Test (1976) as a dependent variable. The average effect size was .59 although the Glassberg study (1980) showed no change in ego judgment. The Loevinger test assesses judgments about interpersonal and intrapersonal relation. Loevinger defines stages of ego as encompassing self-understanding and social perspective taking. The self (ego) is capable of increasingly more complex perspectives on self and self-in-social relationships. Each higher stage

Table 1

Summary of Effects Across Three Professional Judgment Domains for DPPE Studies in Teacher Education and Teacher Professional Development

Career Phase and Study		Rest DIT (P score) Posttest			Loevinger Posttest			Hunt CL or CCI Posttest		
		Mean	*SD*	Effect	Mean	*SD*	Effect	Mean	*SD*	Effect
Pre-Service Teacher Education										
Watson & Reiman (2001)	E	43.33		P				1.94		CL
(Early Field, N = 39)	C	23.83	20.9	+.617				1.74	.43	+.628
Hsiang (1999)	E	42.91		P						
(Early Field, *N* = 51)	C1	35.07	14.87	-.52						
	C2	41.66	11.57	-.10						
Mann (1992*)	E							378.82		CCI
(Tutoring, *N* = 29)	C							367	37.82	+.31
Reiman & Panamone (1993)	E	56.9		P				1.9		
(Tutoring, *N* = 60)	C	46.4	13A	+.93				2.0	.61	-.16
Glassberg (1977)	E	62.0		P	7.34		SCT			
(Student Teaching, *N* = 30)	C	60.29	NA	.00	6.71	NA	.00			
Oja & Sprinthall (1978)**	E	63.1		P	6.90		SCT	2.01		CL
(Student Teaching, *N* = 55)	C	51.5	1.372	+.845	6.56	.821	+.414	1.787	.322	+.692
Riggsbee (1995)	E	60.3		P				1.85		CL
(Student Teachers, *N* = 24)	C	35.1	NA	0.0				1.74	NA	0.0

Table 1 (continued)

Summary of Effects Across Three Professional Judgement Domains for DPPE Studies in Teacher Education and Teacher Professional Development

		Rest DIT (P score) Posttest			Loevinger Posttest			Hunt CL or CCI Posttest		
		Mean	*SD*	Effect	Mean	*SD*	Effect	Mean	*SD*	Effect
Induction-First Three Years of Teaching										
Reiman & Watson (1999) (Beginning Teachers, *N*=9)	E	32.84		P				1.55		CL
		40.00	8.66	+.82				1.90	.31	+1.12
Ongoing Professional Development										
Exum (1980)	E				7.06		SCT			
	C				5.83	.98	+1.37			
Reiman (1988) (Mentoring, *N*=34)	E	48		P				1.96		CL
	C	42	15.86	+.37				1.88	.324	+.22
Parker (1994)** (Middle School Assistance Teams, *N*=16)	E	45.67		P				1.96		CL
	C	30.7	9.87	+1.51				1.68	.32	+.75
Reiman & Peace* (2002) (Peer Coaching and Assistance, *N*=12)	E	52.8		P				2.19		CL
	C	44.3	3.09	+3.0				2.06	.15	+.93
Average Effect Size		(*N*=10)		+.75	(*N*=3)		+.59	(*N*=9)		+.50
		MJI or DIT			SCT			CL or CCI		

Note. One-half of the studies were quasi-experimental designs. A Meta-analysis (*N*=500) studies by Cohen, Kulik, and Kulik (1982) found no differences in effect size by quasi versus "true" experimental designs.

consists of a more differentiated and integrated social perspective.

Nine DPPE studies assessed teachers' conceptual judgment which is a determination of a person's problem solving judgments and tolerance for ambiguity in complex "human-helping" decision-making. The test has a scoring scheme from 0 to 3.0. The Hunt scheme is helpful to our understanding of educators because it also identifies a person's ability to read and flex with learners. Those persons in the 2.0-3.0 range exhibit greater ability to read and flex to learners than teachers in the 1.0-1.9 range. Reading means that the educator is sensitive to the learners' verbal and non-verbal cues, whereas flexing means that those persons who are at higher stages exhibit an ability to flexibly adapt their instruction to better meet pupil needs. The Hunt instrument also can be used to assess educator's reflective thinking – that is, how person's formulated judgments about the real world problems they were facing in classrooms and mentoring situations. The average effect reported in Table 1 is 50.

Changes in Performance

The DPPEs also were designed to foster transformations in performance of prospective teachers and experienced teachers. Although this data is not reported in Table 1, significant transformations were observed. For example, Mann (1992) analyzed audiotapes of tutoring conferences to assess listening performance of tutors. Significant differences were found between the comparison group of tutors and the experimental group that participated in the DPPE. Likewise, Glassberg (1977) found significant positive gains in student teacher performance in engaging learners and responding to learner concerns ($p < .01$). Reiman and Watson (1999) investigated changes in performance of beginning teachers. They found significant changes ($p < .01$) in beginning teacher "motivating interactions" with students.

Summary of Findings

We have already noted that Miller found significant trends between conceptual judgment and teacher openness to student perspectives. Teaching professionals who process experience more complexly in the conceptual judgment domain have a greater ability to "read and flex" with learners, to take the emotional perspective (empathy) of others, and to think on their feet and find alternative solutions (less "functional fixedness") (O'Keefe & Johnston, 1989).

It should be noted that there are equally important relationships between moral/ethical judgment and professional action. For example, teachers with more complex levels of moral/ethical judgment hold more humanistic-democratic views of student discipline (Chang, 1994), foster more intellectual and participatory climates in classrooms, and view curriculum and instruction as needing to be adjusted to students' abilities and needs (Johnston, 1989). Further, moral/ethical judgment correlates with measures of civic responsibility ($r = .44$), community involvement ($r = .31$), measures of reflective judgment ($r = .46$), and professional efficacy – valuing of one's occupation ($r = .38$) (Rest, Narvaez, Bebeau, & Thoma, 2002).

The relationships between judgment domains and adoption of more learner-centered instructional strategies have been chronicled by McKibbin and Joyce (1981) and Hopkins (1990). In the yearlong McKibbin and Joyce study, they demonstrated strong relationships between teacher judgment levels (i.e., psychological states) and effective implementation of innovative instructional methods focusing on engaging students. Teachers at less complex "states" virtually closed their doors to innovations, whereas teachers at more complex judgment states understood the need for change and the need to be responsive to learners' point of view. Hopkins corroborated these findings.

In summary, moderate-to-large effects for the three professional judgment domains were found. These findings confirm that the deliberate psychological and professional education interventions (DPPE) promoted positive and significant transformations in disposition as defined in this paper. In addition, there was consistent evidence that DPPEs fostered transformations in professional performance.

The summarized studies (see Table 1) suggest that the DPPE framework promotes significant changes in professional judgment and professional action. In each case, the studies asked the individual to assume a complex new role that includes guided inquiry, a scope-and sequenced-curriculum, support, and challenges to enlarge both his or her social perspective and socio-moral commitments, while simultaneously acquiring new abilities and new knowledge (i.e., reflective coaching).

Implications for Teaching, Teacher Education, and Research

We have proposed a promising new direction in the study of dispositions in teacher education. Interest in assessing disposition needs to be linked to concepts of teacher professional interpretation, judgment and action. Such scholarship could resolve the persistent problem of lack of clarity in definitions of disposition.

We also noted in our statement of the problem that there is a general lack of evidence regarding the relationship between dispositions and teacher behavior. In general, most researchers would agree that a failure to support such a link between judgment and action would significantly diminish the status of the proposed inquiry. Current evidence on the links between conceptual judgment and teacher action, and moral/ethical judgment and teacher action are promising. It finds that teachers at more complex judgment levels across the ethical and conceptual/reflective domains are more learner centered, more flexible, and more interested in encouraging discourse 2 classroom discussions which are inductive and student-centered. It is also the case that much work remains in charting how teachers' understandings and

judgments about particular situations relate to his or her responsible actions and to student learning outcomes.

Yet Murray's (1996) review illustrates just how far we have to go. It seems very far indeed from current debates about standard-based reform. Linking the national call for assessment of dispositions to current and future scholarship on domains of teacher professional judgment and responsible (e.g., responsive) action has potential. Teacher education programs could initiate pre-and post- assessments of teacher candidates, and begin to develop a data base on how programs foster changes in professional interpretation, judgment, motivation and action. Additional considerations may include evaluation by qualitative means such as observing professional actions and cycles of development over time. Connections between such descriptive methods and quantitative analysis can perhaps be integrated into assessment procedures for teacher education programs.

A final implication of this work relates to the potential of DPPE as a promising theoretical and conceptual model of intervention for teacher education and teacher professional development programs. One of the long-standing problems in this area has been the lack of directing constructs and supporting evidence. Providing opportunities for complex new roles for educators represents a powerful professional development activity that fosters significant transformations in teachers' professional judgment and actions. The researched DPPE programs that are described in the research review fostered significant changes in teachers' conceptual, ethical, and ego levels. The effect sizes for these changes ranged from +.50 to +.75. Further, these studies represent the career span of teacher professional development – preservice – induction – ongoing professional development. The interventions encouraged teachers to reflect on their performance, their views about their work, their beliefs about what learning is, and what constitutes good teaching. The whole professional repertoire is involved when fundamental changes in the teacher role take place. The complex new role requires a reconstruction of basic views on learning and teaching. New teaching roles encompass a range of innovations entailing new goals, new structures, and new behaviors. Designing and implementing DPPEs requires a new conceptualization of professional development and a commitment to a theory of assistance to guide program design. Professional development programs could be organized around the seven essential elements of the DPPE: contextualized learning and development, new role-taking, guided inquiry, balance, continuity, support and challenge, and reflective coaching . And it is supported by continuing research on teacher professional development and its relationship to curriculum and student learning (Ball & Cohen, 1999; Fink & Resnick, 2001; Guskey, 2002; Hargreaves & Fullan, 2000; Kennedy, 1998).

Comprehensive teacher development takes time and assumes that change must move in the direction identified as mature, adaptive, increasingly complex, differentiated, and well-integrated.

References

Andrew, M. D., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of an eleven-university consortium. *Action in Teacher Education, 17*(3), 43-53.

Association of Supervision and Curriculum Development (1962). *Perceiving, behaving, becoming: A new focus for education*. Washington, DC: Author.

Ball, D. L. & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional development. In L. Darling-Hammond and G. Sykes (Eds.), *Teaching as a learning profession* (pp. 3-32). San Francisco: Jossey-Bass.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

Bowen, H. W. (1977). *Investment in learning: The individual and social value of American higher education*. San Francisco: Jossey-Bass.

Chang, F-Y. (1994). School teachers' moral reasoning. In J. Rest and D. Narvarez (eds.), *Moral development in the professions* (pp. 71-84). Hillsdale, NJ: Erlbaum.

Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.

Collinson, V., Killeavy, M., & Stephenson, H. (1999, October). Exemplary teachers: Practicing an ethic of care in England, Ireland, and the United States. *Journal for a Just and Caring Education, 5*(4), 340-366.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Dewey, J. (1904). The relation of theory to practice in education. In *The Third Yearbook of the National Society for the Scientific Study of Education: Part I: The relation of theory to practice in the education of teachers* (pp. 9-30). Chicago: University of Chicago Press.

Dewey, J. (1938). *Experience and education*. New York: Collier Macmillan.

Exum, H. (1980). Ego development: Using curriculum to promote growth. *Character Potential, 9*, 121-128.

Feiman-Nemser, S., & Remillard, J. (1996). Perspectives on learning to teach. In F. Murray (Ed.), *The teacher educator's handbook: Building a knowledge base for the preparation of teachers* (63-91). San Francisco: Jossey-Bass.

Fink, E. & Resnick, L. (2001). Developing principals as instructional leaders. *Phi Delta Kappan, 82*(8), 598-606.

Gage, N. (1985). *Hard gains in the soft sciences: The case for pedagogy*. Bloomington, Indiana: Phi Delta Kappa.

Glassberg, S. (1977). *Peer supervisor for student teachers: A cognitive-developmental approach to teacher education*. Unpublished doctoral dissertation, University of Minnesota.

Glassberg, S. & Sprinthall, N. (1980). Student teaching: A developmental approach. *Journal of Teacher Education, 31*(2), 31-35.

Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership, 59*(6), 45-51.

Hargreaves, A. & Fullan, M. (2000). Mentoring in the new millennium. *Theory into Practice, 39*(1), 50-56.

Holmes Group (1995). *Tomorrow's schools of education: A report of the Holmes Group.* East Lansing, MI: Author.

Hopkins, D. (1991). Integrating staff development and school improvement: A study of personality and school climate. In B. Joyce (Ed.), *ASCD Yearbook: Changing school culture through staff development* (pp. 41-67). Alexandria, VA: Association for Supervision and Curriculum Development.

Hsiang, M. (1999). *Technology-assisted reflection: A study of pre-service teacher education in middle school language arts and social studies and secondary English education and social studies.* Unpublished doctoral dissertation, N.C. State University, Raleigh, N.C.

Hunt, D. (1974). *Matching models in education.* Toronto, Canada: Ontario Institute for Studies in Education.

Interstate New Teacher Assessment and Support Consortium (1992). *Model standards for beginning teacher licensing, assessment, and development: A resource for state dialogue.* Washington, D.C.: Author.

Johnston, M. (1989). Moral reasoning and teachers' understanding of individualized instruction. *Journal of Moral Education, 18*(1), 45-59.

Johnston, M., & Lubomudrov, C. (1987). Teachers' level of moral reasoning and their understanding of classroom rules and roles. *Elementary School Journal, 88*, 65-78.

Joyce, B. & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal.* (2nd Ed). White Plains, NY: Longman.

Katz, L., & Raths, J. (1985). Dispositions as goals for teacher education. *Teaching and Teacher Education, 1*, 301-307.

Kennedy, M.M. (1997). *Defining an ideal teacher education program (mimeo).* Washington, DC: National Council for the Accreditation of Teacher Education.

Kennedy, M. M. (1998). *Form and substance in in-service teacher education* (Research Monograph No. 13). Arlington, VA: National Science Foundation.

Leming, J. (1992). The influence of contemporary issues curricula on school-aged youth. *Review of Research in Education, 18*, 111-161.

Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 48*, 1181-1209.

Loevinger, J. (1976). *Ego development.* San Francisco: Jossey-Bass.

Mann, A. (1992). *A quantitative and qualitative evaluation of a peer tutor-training course: A cognitive-developmental model.* Unpublished doctoral dissertation, N.C. State University, Raleigh.

McKibbin, M., & Joyce, B. (1981). Psychological states. *Theory into Practice, 19*, 248-255.

Mead, G.H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Mentkowski, M., and Associates (2000). *Learning that lasts.* San Francisco: Jossey-Bass.

Miller, A. (1981). Conceptual matching models and interactional research in education. *Review of Educational Research, 51*(1), 33-84.

Mosher, R., & Sprinthall, N. (1970) Psychological education in secondary schools. *American Psychologist, 25*, 911-24.

Murray, F. (1996). Beyond natural teaching: The case for professional education. In F. Murray (Ed.), *The teacher educator's handbook: Building a knowledge base for the preparation of teachers* (pp. 3-13). San Francisco: Jossey-Bass.

National Board for Professional Teaching Standards (1998). *Policy position.* http://www.nbpts.org/nbpts/standards/policy.html.

National Council for the Accreditation of Teacher Educators (2000). *The standard of excellence in teacher preparation.* http://www.ncate.org/.

Oja, S.N., & Smulyan, L. (1989). *Collaborative action research: A Developmental Process.* London: Falmer Press.

Oja, S., & Sprinthall, N.A. (1978). Psychological and moral development for Teachers: Can you teach old dogs? In N. Sprinthall and R. Mosher (Eds.), *Value development as the aim of education*, (pp. 117-134). NY: Character Research Press.

O'Keefe, P,. & Johnston, M. (1989). Perspective-taking and teacher effectiveness: A connecting thread through three developmental literatures. *Journal of Teacher Education, 40*, 14-20.

Oser, F. (1992). Morality as professional action: A discourse approach for teaching. In F. K. Oser, A. Dick, & Patry, J. (Eds.), *Effective and responsible teaching: The new synthesis* (pp. 109-125). San Francisco: Jossey-Bass.

Oser, F., Dick, A., & Patry, J. (Eds.) (1992). Responsibility, effectiveness, and the domains of educational research. In F. Oser, A. Dick, and J. Patry (Eds.) *Effective and responsible teaching: The new synthesis* (pp. 3-13). San Francisco: Jossey-Bass.

Parker, T. (1994). *The effects of a staff development program based on a cognitive-developmental framework on middle school pre-referral assistance teams.* North Carolina State University, Raleigh.

Raths, J. (2000). Teacher beliefs and teaching beliefs. *Early Childhood Research & Practice (18)*, 385- 391.

Raths, L. E., Harmin, M., & Simon, S. B. (1966). *Values and teaching.* Columbus, OH: Merrill.

Reiman, A. J. (1999). The evolution of the social role-taking and guided reflection framework in teacher education:

Recent theory and quantitative synthesis of research. *Teaching and Teacher Education, 15*, 597-612.

Reiman, A. J. & Peace, S. D. (2002). Promoting teachers' moral reasoning and collaborative inquiry performance: A developmental role-taking and guided inquiry study. *Journal of Moral Education, 31*(1), 51-66.

Reiman, A. J., Sprinthall, N. A., & Thies-Sprinthall, L. (1999). The conceptual and ethical development of teachers. In R. Mosher, D. Youngman, & J. Day (Eds.), *Human development across the life span: Educational and psychological applications* (pp.203-218). Westport, CT: Praeger.

Reiman, A. J., & Watson, B. (1999). Beginning teacher development: A cognitive-developmental approach. *Journal of Research in Education, 9*(1), 1-10.

Reiman, A. J., & Thies-Sprinthall, L. (1998). *Mentoring and supervision for teacher development.* New York: Addison-Wesley Longman.

Reiman, A. J., & Parramore, B. (1993). Promoting preservice teacher development through extended field experience. In M. O'Hair and S. Odell (Eds.), *Teacher Education Yearbook I.* Fort Worth, TX: Harcourt Brace Jovanovich.

Rest, J. (1986). *Moral development.* New York: Praeger.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Riggsbee, J. (1995). *A developmental reflective seminar for student teachers: A cognitive developmental approach to teacher education.* Unpublished dissertation. Raleigh, NC: North Carolina State University.

Rosenthal, R. (1995). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Selman, R. (1980). *The growth of interpersonal understanding.* Orlando: FL: Academic Press.

Selman, R., Watts, C. L., & Schultz, L. H. (Eds.) (1997). *Fostering friendship.* New York: Aldine de Gruyter.

Shulman, L. S. (1998). Theory, practice, and the education of professionals. *The Elementary School Journal, 98*(5), 511-526.

Sikula, J. (Ed.) (1996). *Second handbook of research on teacher education.* New York: Macmillan.

Sprinthall, N. A., & Thies-Sprinthall, L. (1983). The teacher as an adult learner. In G. A. Griffin (Ed.), *Staff Development: Eighty-second Yearbook of the National Society for the Study of Education* (pp. 24-31). Chicago: University of Chicago Press.

Sprinthall, N. A., Reiman, A. J., & Thies-Sprinthall, L. (1996). Teacher professional development. In J. Sikula (ed.), *Second handbook of research on teacher education.* NY: Macmillan.

Tatto, M. T. (1996). Examining values and beliefs about teaching diverse students: Understanding the challenge for teacher education. *Educational Evaluation and Policy Analysis, 18*(2), 155-180.

Taylor, R. D., & Wasicsko, M. M. (2000, November). *The dispositions to teach.* Paper presented at the Southern Regional Association of Teacher Educators, Lexington, KY.

Thies-Sprinthall, L. (1984). Promoting the developmental growth of supervising teachers: theory, research programs, and implications. *Journal of Teacher Education, 35*, 13-20.

Vygotsky, L. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Watson, B., & Reiman, A. (2001). Promoting ethical and reflective professional judgment in pre-service teachers. In J. Ranier and E. Guyton (Eds.), *Research on the effects of teacher education on teacher performance* (pp. 29-44). Dubuque, IA: Kendall/Hunt Publishing.

Alan Reiman is an Associate Professor in the Department of Curriculum and Instruction at North Carolina State University, Raleigh. He also is Executive Director of SUCCEED at Mentoring. His research and writing focuses on theory, research, and practice for Deliberative Psychological and Professional Development programs in teacher education and staff development.

Lisa Johnson is a National Board Certified teacher and doctoral student in the Department of Curriculum and Instruction at North Carolina State University, Raleigh. Her research focuses on the relationships between ethical and reflective judgment and professional action in teacher mentoring dyads.

Children's Perceptions of High Stakes Testing

Cheri Foster Triplett
and
Mary Alice Barksdale
Virginia Tech

Paula Leftwich
University of South Florida

The perceptions of students reported in this manuscript come from a larger study that was designed to explore children's perceptions of schooling in general. In this manuscript, we focus on the theme of high stakes testing which was a salient issue in the larger study. Although there were no explicit questions about high stakes testing in our interviewing protocol, 17 of the original 33 students expressed concerns about high stakes testing. These students reported numerous emotional and physical responses to test taking, including worrying and losing sleep. Also students reported social concerns about failure, grade retention, and their personal reputations. Through these interviews, we provide a glimpse at the stakes according to children.

The social and emotional aspects of children's lives have an impact on their progress in school; in fact, in schools where clear efforts are made to address social and emotional development, academic development is enhanced (Darling-Hammond, 1996; Green, 1998; Kaplan & Evans, 1997; Noddings, 1992, 1995). However, there are indications that, in general, children are feeling increasingly anxious about schooling and their performance (Barkdale, McGill-Franzen, & Leftwich, 2002). Likewise, recent research revealed that high stakes testing contributes to children's anxiety (Barksdale-Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001).

Children in the United States are provided instruction that is increasingly centered on test preparations. Although our children come from very diverse families, regions, and cultures, at school, they are expected to meet common standards. During the nineties, educators in almost every state created descriptions of standards at specific grade levels and in various content areas (McLaughlin, 1994). Established as state education policies, standards have been placed in the hands of teachers while simultaneously, millions of dollars have been spent on the development of tests designed to measure student performance on each standard. There are broad differences in the tests developed in different states, with some being identified as norm-referenced, some as criterion-referenced, and some as performance assessments (Haertel, 1999; Sacks, 1997; Sheldon & Biddle, 1998). Thus, our nation's children are expected to demonstrate, through scores on tests, that they have met specific standards, and in some cases, there are serious consequences for children if they do not achieve minimum scores (Chase, 2000). Furthermore, research (Hoffman, Assaf, & Paris, 2001; Gordon & Reese, 1997; Langenfeld, Thurlow, & Scott, 1997) showed that children considered to be at-risk experience some of the most serious consequences, including overall effects on self-esteem.

The consequences for teachers have also been well documented by researchers (Gordon & Reese, 1997; Johnston, Afflerbach, & Weiss, 1993; Johnston, Guice, Baker, Malone, & Michelson, 1995; Smith, Edelsky, Draper, Rottenberg, & Cherland, 1991). The publishing of test scores, the grading of schools in local newspapers, and the pressure to produce high test scores has caused teachers to suffer from "anxiety, shame, loss of esteem, and alienation" (Smith, 1991, p. 8). When the stakes are high for teachers, they tend to focus on preparing for the test. When there is a major instructional focus on preparing children for tests, teachers move toward teaching specifically to the test while ignoring other aspects of instruction (Hoffman, Assaf, & Paris, 2001; Johnston, 1998; Herman & Golan, 1991; Smith, 1991). Several studies revealed that teachers made significant changes in classroom instruction as a result of increased pressure to assure that children have high test scores. Teachers reported discontinuing activities that are hands-on, inquiry based, collaborative, and integrated in favor of rote level, discrete, individual drill and skill practice (Barksdale-Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001; McMillan, Myran, & Workman, 1999).

No doubt, these changes in instruction affect students, but students also experience anxiety surrounding test taking itself. Although we have few self reports from students, teachers reported emotional and physical effects on students,

including anxiety, panic, irritability, frustration, boredom, crying, headaches, and loss of sleep (Barksdale-Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001). In Hargrove, Jones, Jones, Hardin, Chapman, & Davis' (2000) research surveying teachers' perceptions of high stakes testing, teachers reported an overall effect on students' morale and love of learning.

Although researchers have queried the perceptions of teachers (Barksdale-Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001), parents (Barksdale-Ladd & Thomas, 2000), administrators (Nolen, Haldyna, & Haas, 1989), student teachers (Sturtevant,White, & Dunlap, 2001), and even policy makers (Atkinson & Miller, 1999), little is known about children's perceptions of high stakes testing. In the midst of nationwide attention to test scores, individual children have not been heard — little is known about their thoughts and feelings related to testing.

The perceptions of children reported in this manuscript come from a larger study that was designed to examine the lived experience of elementary and adolescent children in public schools. Specifically, the purpose of the study was to explore (a) student perceptions of their experiences in school, (b) areas of schooling viewed as most successful by students and what aspects of student lives support these successes, (c) areas of schooling found to be most difficult for students and how students deal with or seek support in these problematic areas, and (d) what stresses are experienced by students and how students deal with the stresses in their lives.

In this manuscript we focus on the theme of high stakes testing, which was a salient category from the larger study. Although children were not asked explicitly about testing, over half (17) the students expressed concerns about testing, thus we wanted to look in greater detail at the issues children raised about testing.

Methods

Participants

The participants in this study were 33 third through eighth-grade students. The students attended our university partnership schools. Relationships with principals and teachers at these schools were longstanding through these partnerships. Likewise, some teachers and principals were enrolled in our university graduate programs in education. Each researcher visited several classrooms to describe the study and invite students to participate. All students who participated were those who volunteered and those whose parents provided permission for participation in the study. Thus, a total of 33 students from five different classrooms volunteered for this study. The students chose their own pseudonyms, which were used exclusively during the interviews. No records were kept of the real names; thus, student-selected pseudonyms have been used in reporting the results.

In one state, although students take a state-mandated test in grades 3-10; grades 4, 5, 8, and 10 are key levels in which test results are regarded as particularly high-stakes. Students in Grades 4, 8, and 10 have in common the emphasis on literacy assessment, both in Reading and Writing. Schools receive a letter grade based largely on test results. Financial rewards and recognition accrue to schools where the grade is high, and the threat of closure and reconstitution faces schools with persistent low grades. Pressure to produce satisfactorily high test scores is intense and pervasive for students, teachers, and administrators, especially in schools where past test scores and schools grades have been low.

Eighteen elementary students in this state were interviewed during the spring of 2001. In one school, in a rural area, ten students from one fourth-grade classroom were interviewed (Jaguar, Brandon, Meshell, Rocky, Whitney, Iddy Bug, Puppy, Dobby, Jasmine, and Harry). Each of these students was interviewed three times. Interviews were conducted prior to testing, soon thereafter, and about one month later. In an urban school in this same state, eight students from one fifth-grade classroom were interviewed (Princess, Aslin, JJ, JC, Gohan, Gem, Suzy, and Symone). These students were interviewed one time, soon after taking the state-mandated test. The two classrooms were chosen for their composite student list that included a representative mix of students in terms of gender, race, SES, and achievement.

In this same state, eighth-grade students (Dominic, Angelica, Todd, Aimee, Zack, Michael, Maria, and Celeste) from a medium-sized middle school in a rural-suburban public school district were interviewed in mid-spring of 2001. The classes were chosen for their composite student list that included a representative mix of students in terms of gender, race, SES, and achievement.

In a second state, where the testing milieu is increasingly high stakes, all third graders take a state mandated test that covers reading, writing, mathematics, history and social science, and science. Seven third-grade students were interviewed during the spring of 2001. Each student was interviewed three times. Interviews were conducted prior to testing, soon thereafter, and about one month later. These students were selected from two classrooms, one in a small town (Pokie, Brittany, Ashley, and Danielle), and the other in a suburban area (Wolf, Sissy, and Nick).

Research Design

The research design was qualitative, with interviews serving as the primary data sources. Interviews were semi-structured, using the following interview protocol:

1. Tell me about how things are going for you at school right now.
2. Tell me about what you are learning right now.
3. Is there anything happening in school right now that is particularly helpful to you in learning what your teachers expect you to learn?

what your teachers expect you to learn?
4. Are there any areas in which you are being particularly successful right now? (If so,) tell me about your success. What are you doing that is helping you to be successful in this area?
5. Are there any areas in which you are struggling, or being less successful right now? (If so,) tell me about the difficulty you are having. Is there something you should be doing that would help solve the problem? Is there something that someone else could do to help you with this difficulty?
6. Are there any aspects of your life at school that are stressful for you right now? (If so,) Tell me about these stresses.
7. Are there any aspects of your life away from school that are stressful for you right now? (If so,) tell me about these stresses.
8. (If stresses were identified in # 6 or # 7, What is happening that helps you deal with the stresses in your life right now? Is there anything you are doing that helps you deal with the stresses in your life right now?
9. Is there anything else that you would like to tell me about today?

Probing statements were used to follow-up comments made by the students. Probes included statements such as: (a) "Tell me more about that," (b) "I would like for you to explain that for me a little further," and (c) "Is there anything you'd like to add?"

The interview protocol included *no* questions about testing. We wanted to discover whether or not the participants would discuss testing. If the participants brought up the topic, we generally asked some probing questions; however, we were careful to limit these inquiries. We did not wish to provide feedback to the students indicating that we had a particular interest in testing above other topics, as this could have influenced future interviews. Interviews were held in private areas in the schools where other students could not overhear the participants. Interviews lasted about 20 minutes on the average. We listened for as long as the participants wanted to talk, and the longest interviews were of about 35 minutes.

Our careful interviewing protocol, in a sense, became problematic. We sought to maintain consistency across the interviews, and yet the information that we were able to gather was then limited by this procedure. For instance, when students did not mention testing, we were not able to proceed with that line of inquiry. Interviewing also became a logistical conundrum, finding a private place to interview was often difficult. We recognized that students were not as vocal about their thoughts and feelings when our context was merely semi-private, i.e. the back of a classroom.

Data Analysis

Each interview was audio taped. Interviews were transcribed and read by members of the research team. Interviews were segmented into units of meaning and careful attention was given to the preservation of context within the units of meaning (Hycner, 1985). After segmenting the interviews, the units of meaning were grouped into categories sharing common themes or characteristics. After organizing these data into categories, each category was given a name (or code) that was representative of commonalties within the grouped units of meaning (Bogdan & Biklen, 1998). Units of meaning were double-coded if they contained information relative to multiple categories. We focus on one of those coded categories in this paper — high stakes testing.

Findings and Discussion

The data representative of high stakes testing emerged from responses to questions (5) "Are there any areas in which you are struggling, or being less successful right now?" and (6) "Are there any aspects of your life at school that are stressful for you right now?" However, as mentioned previously, students were not asked directly about testing. As we perused the data representative of high stakes testing, we were able to discern four specific patterns or subcategories that were salient: (a) who was worried about high stakes testing? (b) when did students worry? (c) what were the emotional and physical effects of test taking? and (d) what were the stakes according to children? Consulting previous research in high stakes testing likewise revealed that these particular patterns were worth exploring.

Who was worried about high stakes testing?

Amongst the students interviewed, there were some interesting differences in who reported concerns about high stakes testing. None of the children from the urban school mentioned testing as a worry. Actually, none of the children from the urban school mentioned testing at all. This was a surprising finding, given that the urban students in two previous studies (Wheelock, Bebell, & Haney, 2002; Young, 1994) reported more concerns about testing than their suburban peers. The children from the urban school were interviewed only once, but it was soon after they had taken their state-mandated test. Because they did not bring up the topic of testing, they were not asked any questions about it. The research design didn't provide for an investigation into why the students in this one school did not worry about testing — this was certainly a limitation in our interviewing protocol. Our best explanation is that this urban school was relatively new and the attention of the principal, faculty, students and parents was focused on becoming a cohesive community instead of being focused on testing.

Likewise, none of the eighth-grade students expressed concerns about testing, their worries and concerns centered on their peer relationships. This is not surprising given that middle schoolers may develop a disillusionment about standardized testing and it's importance (Paris, Turner, Lawton, & Roth, 1991; Wheelock, Bebell, & Haney, 2002)

and may not report worries about schooling in general Barksdale, McGill-Franzen, & Leftwich, 2002; Nichols & Hancock, 2002).

In the other three elementary schools, both rural and suburban, all of the students discussed worries about testing:

> Well, we HAVE to take 'um...so it makes me worry a lot before they come, and then you've got to worry after that to see if you passed. (Puppy)

> I am worried about the FCAT. When I was in Mrs. (teacher)'s class doing my FCAT, I was worried I will not get a good grade on it because there are a lot of the problems that I don't know and it was like, think, think, think. And I would write it, and that's what I'm worried about. I don't want to get a bad grade on that. (Jaguar)

> I worry about it, like, for the test, like, what if I fail and everybody else gets a good grade. I sort of worry about that. And like, hard things. I try to think and think, but I still don't know what the answer is, and that worries me a lot. I get all nervous because there is like one bubble that you got to fill in. It was really hard, most of the tests that we had to bubble in. (Whitney)

One third grader, Nick, even reported worrying about making his parents worry. This child reported being "very nervous and scared" before taking the test and that he "worried a lot about this every day." However, he did not want to tell his parents or his teachers that he was worried, because "they might get worried about (him) and they might start getting a bit nervous."

When did students worry?

Students reported that their teachers worried "all year long" about testing. According to Wolf, "The teacher talked about it every day the whole year...I just knew, I could tell Ms. (teacher) was worried 'cause everything we did, every single day, she would be saying something about remembering it for the SOL test." Other students also reported that the year long test preparations caused them anxiety:

> ***Sissy***- I was always nervous about it because we studied and studied and practiced a whole lot in school.
> [How did you practice?]
> ***Sissy***- We were practicing everything, like writing little stories like you have to do on the SOLS and doing the same kind of math problems and all of that.
> [Did your teacher tell you that it was practice for the SOL test?]
> ***Sissy***- Oh yes, Ms. (teacher) told us every time, and I think maybe we did a little practice every day.
> [All year?]
> ***Sissy*** -Yeah, but after you take it, you don't have to practice any more.

In previous studies (Barksdale -Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001) teachers reported their year long anxiety over test preparations, including "bubbling" practice, test taking skills, content reviews, and practice tests.

While some students reported worrying during test preparations, others reported worrying during the testing experience or afterward, waiting for their scores. Some students, in both states, reported worrying before, during, and after testing. Dobby is a prime example:

> [Tell me about your worries about these tests.]
> ***Dobby*** -It's pretty scary, so you can't help but worry.
> [Can you tell me more about that?]
> ***Dobby*** - It's just that the teacher keeps getting ready for you to take the test, practicin and stuff, and you keep on getting ready, and while you are doing that, you get worrying about it...and it's a really hard test, and long, it takes a long time to do it, with hundreds and hundreds of questions, and bubble sheets to answer, and it's hard to stay on the right line, like, very confusing, and you have to get almost all the questions right or else it will be failing, and if you fail the FCAT, everyone will be mad, like your teachers and Mommy and Daddy, and you might fail your grade and have to to it all over again.

These elementary students not only reported worrying before, during and after test taking, they reported many other emotional and physical responses.

What are the emotional and physical effects of test taking?

Similar to the reports made by teachers in previous studies (Barksdale-Ladd & Thomas, 2000; Gordon & Reese, 1997; Hoffman, Assaf, & Paris, 2001), children's self reports in this study included numerous emotional effects such as crying and nervousness.

> I wanted to cry. Ms. (teacher) gave it out, and that sheet was there, the one with bubbles, and I felt like crying. I couldn't think and I couldn't remember and I think probably, I got an F on it because I just couldn't do that stuff, and for so long. I don't think, I mean, probably, we are just little kids and we shouldn't oughta probably have to do something like this, but they make us. (Jasmine)

> There was a part where you had to explain it, you

had to describe something, and I was really nervous. I kept on thinking I was going to mess up...and I kept on saying I was going to get it wrong and I don't know if I did or not. (Brittany)

These findings concur with the reports of teachers — some children experience a lot of anxiety over testing. Also similar to the reports made by teachers in previous studies (Barksdale-Ladd & Thomas, 2000; Hoffman, Assaf, & Paris, 2001), students in this study reported numerous physical consequences due to this test anxiety.

It was hard and it made you nervous. And it made you tired because you got to get up early to study. I had to. I don't know about other people. And you needed a drink of water. And you were hungry...you had to be there a very long time. (Meshell)

I was pretty nervous when the FCAT started" [Tell me more about feeling nervous] "Well what goes to my head when I am nervous... I am like sweating and stuff, my brain says to try to do better and you don't got to get all the answers right...so sometimes I just guess because it is a little hard...so when I get nervous sometimes I sweat...I just most of the time sweat. (Brandon)

I was nervous at night, and you know, like when we took it we were supposed to go to bed early, but I couldn't fall asleep cause I was still thinking about it. (Sissy)

The SOL test, that is the worst thing that happens in school. It is awful and I worried and worried and my mom and dad, they worried a lot and we talked about it every day...I thought I would probably throw up in the middle of the test, it just seemed like it made me sick... (Wolf)

Although it is unclear how emotional and physical effects influenced these children's testing performance, the negative effects of anxiety on test performance are well documented (Hembree, 1988; Wigfield & Eccles, 1989).

What are the stakes according to children?

Beyond the emotional and physical consequences surrounding test taking, the children in this study understood the stakes associated with passing or failing the required test. The majority of students were concerned about going to summer school or failing their grade. Rocky was concerned that his FCAT failure would result in going to summer school. He reported worrying about going to summer school "every day." Brandon was happy that he passed the test and could move to fourth grade with his friends. He commented, "I past my FCAT test so I get to go up to fourth grade next year." When asked to explain, Brandon reported, "If you got like all of them right it means you might move up to a higher level next year...if you got like all of them wrong instead of like three or four...they would probably keep you down." Wolf and Brittany also expressed concerns about the consequences of failing the test. Wolf seemed to notice a connection between failing at school and failing the test.

Wolf- I wonder sometimes about other kids because I already get good grades and I know most of the stuff all the time, but those other kids, I mean if I don't know the answers on the test, I know that they don't know even more of them, so what happens to them?
[What do you think happens to them?]
Wolf- It seems like they would fail or have to repeat, I don't know kids who have repeated yet, but some kids in other classes that I don't know, they repeated and it might be because they failed the SOL test.
[Does that worry you?]
Wolf- Yes, yes, I was thinking a lot about it when we took the tests.

Brittany, however, did not see a connection between school performance and testing performance.

It seems like not really fair to me cause you can do good everyday in school and be good and not get in trouble or anything, and if you fail the test, then you are going to probably fail in school, and some of the kids talk about that. (Brittany)

Teachers and parents previously reported their own concerns about the validity of high stakes tests as an indicator of academic achievement and school performance (Barksdale-Ladd & Thomas, 2000; Hoffman, Assaf, & Paris, 2001; Jones, Jones, Hardin, Chapman, Yarbrough, & Davis, 1999). Although it is arguably unscientific to use one test as an indicator of individual student achievement or as a catalyst for retention or promotion, this continues to be a practice in the majority of our nations schools (Allington, 2002).

High stakes testing also made these children feel their reputations were at stake. These elementary students saw high stakes testing as something that can make you "look dumb" to peers and adults.

The FCAT was kind of frustrating. It was, um, kind of hard, and it is going to the state and it's like you don't know who is going to read it, and it is easy to compose a story normally, but some of them for FCAT I had trouble making it up... I had difficulty because you, you like have to read this big story and then you have to answer it, and sometimes I forget and sometimes I skip a sentence, and I am like, I don't remember that. Because any time I skip a sentence, it is always the question... FCAT. I

> scary. It is like, what if I get it wrong and some people think I am stupid. (Iddy Bug)
>
> I'm scared to know if I passed or not, 'cause if not, then everyone will think I'm so dumb. (Wolf)

Whitney reported worrying that her peers would get a good grade on the test while she failed, explaining that this would be very embarrassing. Dobby worried that his parents and teachers would be mad at him if he failed. These students were aware of the social consequences of failing the test. In previous research (Gordon & Reese, 1997), teachers reported that these types of social consquences occur mostly for at-risk students. However, in this study, numerous children, including those who had experienced success in school and those who had not, reported concerns about the social consequences of failure.

Conclusions

In this study, the effects of high stakes testing on children included worry, anxiety, nervousness, sweat, tears, stomach aches, and loss of sleep. These self-reports concur with the previous reports made by teachers. However, teachers have not previously reported the extent to which students worry before, during, and after testing. In this study, elementary students reported worrying before testing, during test preparations. Students detailed the anxiety related to daily practice, including bubbling sheets, writing prompts, math problems and test taking skills. Although teachers in previous research mentioned their own anxiety related to daily drill and skill practice, they did not report children's anxiety during test preparations. Quite possibly teachers are not aware that daily test preparations can cause students to worry or can reveal to students their own worries about testing.

Students also detailed their anxiety during test taking, including worrying about bubbling correctly, worrying about not knowing the answer, worrying about time constraints, and even worrying about whether their peers would pass the test. Likewise many students reported worrying after testing, as they waited for their scores. This study reveals that, in essence, some elementary students worry about their end of grade tests — all year long.

Most importantly, this study details the stakes according to children. These elementary students worried about the consequences of failing the test, including going to summer school and/or failing their grade. Some students in this study, just as some parents and teachers in previous studies, did not think one test should be used to determine retention or promotion. Much like teachers' previous reports about their own fear of social consequences, the elementary students in this study also worried that their reputations were at stake. Would they look dumb? Be embarrassed? Anger their parents? Look stupid to their peers? Disappoint their teachers?

After this research was conducted, the federal No Child Left Behind legislation was passed, requiring even tougher state testing standards and even higher stakes for students who do not pass. Recent reports are grim in states where retention and promotion decisions have been made based on latest test scores. Some states report as many as 40,000 students will be retained because of their spring 2003 test scores. Now that we have grim reports from all the stake holders and research revealing detrimental instructional changes due to high stakes testing, it is incumbent upon us as educators to be advocates in our local and national organizations (Allington, 2002; Hoffman, Assaf, & Paris, 2001). Continued research is our professional responsibility and can be our proactive response to the high stakes testing dilemma.

References

Allington, R. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence.* New York: Heinemann.

Atkinson, T. S. & Miller, S. D. (1999). *Policymakers' perceptions of state education reform: What happens when you swim upstream?* Paper presented at the annual meeting of the National Reading Conference, Orlando.

Barksdale, M. A., McGill-Franzen, A. & Leftwich, P. (2002). *Perceptions of schooling in 4th-8th graders in the United States.* Paper presented at the 19th World Congress on Reading, Edinburgh, Scotland.

Barksdale-Ladd, M. A. & Thomas, K. F. (2000). What's at stake in high stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51,* 384-397.

Bogdan, R. C., & Biklen, S. K., (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Boston: Allyn and Bacon.

Chase, B. (2000). "A sucker punch. Standards, high-stakes tests, and the unlevel playing field." *National Education Association Column,* Jan. 16. http://www.nea.org/publiced/chase/bc000116.html.

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan, 9,* 686-693.

Gordon, S., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership,* 7, 345-368.

Green, R. (1998). The existence of nurturing characteristics in schools and its effect on student discipline, student attendance, and eighth-grade proficiency test scores. *Guidance and Counseling, 13* (4), 10-16.

Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80,* 662-666.

Hargrove, T., Jones, G., Jones, B., Hardin, B., Chapman, L., & Davis, M. (2000). Unintended consequences of high-stakes testing in North Carolina: Teacher Perceptions. *ERS Spectrum, 8* (4), 21-24.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research,*

treatment of test anxiety. *Review of Educational Research, 58*, 47-77.

Herman, J., & Golan, S. (1991). *Effects of standardized testing on teachers and learning - another look. CSE Technical Report No. 334.* Los Angeles: Center for the Study of Evaluation, University of California at Los Angeles.

Hoffman, J., Assaf, L, & Paris, S. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher, 54*, 482-492.

Hycner, R. H. (1985). Some guidelines for the phenomenological analysis of interview data. *Human Studies, 8*, 279-303.

Johnston, P. (1998). The consequences and the use of standardized tests. In S. Murphy (Ed.), *Fragile Evidence. A Critique of Reading Assessment* (pp. 89-101). Mahwah, NJ: Lawrence Erlbaum.

Johnston, P., Afflerbach, P., & Weiss, P. (1993). Teachers' evaluation of the teaching and learning of literacy and literature. *Educational Assessment, 1*, 91-117.

Johnston, P., Guice, S., Baker, K., Malone, J., & Michelson, N. (1995). Assessment of teaching and learning in "literature based" classrooms. *Teaching and Teacher Education, 11*, 359-371.

Jones, M., Jones, B., Hardin, B., Chapman, L., Yarborough, T., Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*, 199-203.

Kaplan, L. S. & Evans, M. W. (1997). Changing school environment: Restructuring one Virginia High School. *NASSP Bulletin, 821*(598), 1-9.

Langenfeld, K. L., Thurlow, M. L., & Scott, D. L. (1996). *High stakes testing for students: Unanswered questions and implications for students with disabilities (Synthesis Report No. 26).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

McLaughlin, M. W. (1994). Educational policy, impact on practice. In M. Aiken (Ed.), *American Educational Research Association Encyclopedia of Educational Research* (pp. 375-382). NY: Macmillan.

McMillan, J. H., Myran, S., & Workman, D. (1999). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Nichols, D., & Hancock, J. (April, 2002). *Anxiety's Influence on German and US Middle Grades Students' Classroom Achievement.* Paper presented at the American Educational Research Association, New Orleans, LA.

Noddings, N. (1995). Teaching themes of caring. *The Educational Digest, 3*, 25-28.

Noddings, N. (1992). *The challenge to care in schools: An alternative approach to education.* New York: Teachers College Press.

Nolen, S., Haladyna, T., & Haas, N. (1989). *A survey of Arizona teachers and school administrators on the uses and effects of standardized achievement testing* (Tech. Rep. No. 89-2). Phoenix, AZ: Arizona State University

Paris, S.,Turner, J., Lawton, T., & Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20* (5), 12-20.

Sacks, P. (1997). Standardized testing: Meritocracy's crooked yardstick. *Change, 29*, 24-28.

Sheldon, K. M., & Biddle, B. J. (1998). Standards, accountability, and school reform: perils and pitfalls. *Teachers College Record, 100*, 164-180.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*, 8-11.

Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1991). *The role of testing in elementary schools. CSE Technical Report No. 321.* Los Angeles: Center for the Study of Evaluation, University of California at Los Angeles.

Sturtevant, E., White, S., & Dunlap, K. L. (2001). *Learning to teach literacy in a high-stakes testing environment: A three-year study of pre-service and beginning early childhood and middle grade teachers.* Paper presented at the National Reading Conference, San Antonio, TX.

Wheelock, A., Bebell, D., & Haney, W. (2002). *What can drawings tell us about high-stakes testing in Massachusetts?* February: http://wwwcsteep.bc.edu/drawoned/mcas/mcaspaper2.html.

Wigfield, A., & Eccles, J. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist, 24*, 159-183.

Young, P. (1994). *Attitudes, perceptions, and beliefs of students and teachers regarding standardized tests.* Unpublished doctoral dissertation, Rutgers University, New Jersey.

Cheri F. Triplett is an Assistant Professor in the Elementary Education program at Virginia Tech. Her research focuses on the socioemotional aspects of literacy learning.

Mary Alice Barksdale is an Associate Professor in the Elementary Education program at Virginia Tech. She is a former Fulbright Scholar and her research centers on teachers and teacher education in literacy, as well as high stakes testing.

Paula Leftwich is an Assistant Professor in the Elementary Education Program at University of South Florida.

Trust in Schools: Extending the Perspective to Institutions

Alan B. Henkin
The University of Iowa

Carole A. Singleton
University of Northern Iowa

Fred B. Holman
St. Thomas University

Jay R. Dee
University of Massachusetts-Boston

Trust can be viewed as a variate within society. Schools as institutions may be conceptualized as objects endowed with trust within the larger social order; trust that may be expected to vary over time. Public trust and confidence in schools may be linked, in part, to trust in these objects as dependable, reliable and effectively functioning educational arrangements. Theoretical and conceptual literatures on trust inform this initial consideration of impersonal dimensions of the phenomenon. Our findings suggest that trust in schools is related to citizen support for schools.

Trust constitutes a vital form of social capital within social systems (Coleman, 1990; Putnam, 1993). "The entire fabric of our day-to-day living, of our social order, rests on trust" (Rotter, 1971, p. 443). Trust is a means for reducing the complexity of society (Luhmann, 1979). It is "a social good to be protected just as much as the air we breathe or the water we drink" (Bok, 1978, p. 26).

Trust is a salient element in constructive human relationships, and is recognized as a critical component of well-functioning organizations in for-profit and not-for-profit sectors (Lane, 1998; Tschannen-Moran & Hoy, 2000). It is a prominent subject of interest in the literatures of school organization and reform (Bryk & Schneider, 2002; Elmore, 1993; Pounder, 1998; Rossow & Zager, 1989), and in research on collaboration and teamwork in schools (Dee & Henkin, 2001; Reyes, 1990). It is recognized, particularly, as an essential element for effective collective action in the context of self-managed schools where there are increased dependencies on teams (Beeson & Matthews, 1993; Henkin & Dee, 2001). "Trust is important because it is a key antecedent of the willingness to cooperate voluntarily. It therefore encourages behaviors that facilitate productive social interaction" (Tyler, 2000, p. 287). Lack of trust in organizations can be costly when employees are unwilling to take risks, demand greater protections against betrayal, and insist on the development and maintenance of expensive and unproductive sanctioning mechanisms to defend their interests (Kramer & Tyler, 1996). Conversely, where social trust prevails in organizations, including schools, people are prepared to take risks, assume that interpersonal and system-level trust is all pervasive, and are willing to invest themselves in groups and teams (Dee & Henkin, 2001). Trust may serve as an enabling force for strengthening both group identity, and for increasing forms of social capital available for the advancement of internally conceived educational purposes (Jones & George, 1998; Kramer & Tyler, 1996). Thereafter, presumptive actions may be predicated on trust of others, and yield high levels of individual satisfaction from successes of the group (Simon, 1991). Such conditions of work can reinforce identification with the collectivity, and provide motivation to work for collective goals (Kramer, Brewer, & Hanna, 1996).

Recent trust research in schools has focused on social relationships between teachers, students, parents and administrators, and has revealed some productivity-related benefits of relational trust that positively impact schools;

specifically, increments in teacher willingness to experiment with new practices, and improvements in teacher-parent relationships, the teaching-learning process, and student outcomes (Bryk & Schneider, 2002). Relevant research conducted in other organizational contexts confirms associations between higher levels of relational trust and outcomes that depend on individual and organizational effectiveness (Butler, 1991; Lane & Bachmann, 1998; McAllister, 1995; Whitener, Brodt, Korsgaard, & Werner, 1998). A recent study in the for-profit sector found, similarly, that three year total return rates to shareholders are almost three times greater in organizations with high trust levels among employees than in companies with relatively low levels of trust (Watson Wyatt, 2002).

From an external perspective, schools as organizations may be viewed as providers of educational services within complex community environments. Communities may include many citizens who are not direct beneficiaries of educational services, and have minimal contact with schools. In such circumstances, schools operate in larger trust markets that may be characterized by asymmetric information (Barber, 1983); school personnel and citizens involved in school activities commonly know more than others about education and community schools. The wider community may rely on assurances that schools will function as expected, and needed services will be delivered. Community perceptions of probity of word and action among successive organizational leaders, over the long term, may strengthen such assurances (LaPorte & Metlay, 1996). Schools may represent a set of core organizational features that have embedded within them the mechanisms required to assure their publics that they can be trusted to provide reliable educational services, and act in the public interest (Creed & Miles, 1996).

Schools are advantaged where communities perceive them as trustworthy. From an institutional perspective, trust production may be viewed as a function of the development and maintenance of organizational settings with structural features and operational conditions that strengthen public confidence, and convey a basic interest in fair dealing. Schools may reinforce citizen perceptions of trustworthiness by placing emphases on those features and conditions that reassure their referent publics.

Schools are attempting to elevate perceptions of trustworthiness, at the same time that they are being subjected to serious political and economic pressures in complex community environments beset with "wicked problems"; problems that are ill defined and rarely, if ever, solved. Every problem can be considered to be a symptom of another problem (Rittel & Webber, 1973). There is reason to suggest, moreover, that the same distrust that may be permeating our society in general may be impacting schools as social institutions (Mathews, 1996).

When a public organization loses "public trust and confidence...many members of the public and stakeholder groups believe the organization neither intends to take their interests into account, nor would it have the competence/ capability to act effectively even if it tried to do so" (LaPorte & Metlay, 1996, p. 342). A disconnect may be occurring between schools and their publics, and the disconnect may be serious. One study suggested that the proportion of Americans indicating confidence in educational institutions from 1973 to 1993, for example, declined from 37% to 22%, at a time when the proportion of those with little confidence rose from 8 to 18% (NORC, 1993). There is no subsequent evidence to suggest that the trend has been reversed. It appears that schools may be viewed, decreasingly, as citizen products of their own creation and, more often, as extensions of the government; further distancing them from their bases of support. Where schools are perceived to be extensions of government, they may be caught up in the trend toward declining confidence in government that has occurred over the last five decades (Purdy, 2003). "Trust in government has fallen by half, since its peak in 1966" (p. 82). Trust in schools may be heading in the same direction.

Perspectives on Institutional Trust

There are many views of trust and many types of trust (Barber, 1983; Dasgupta, 1988; Deutsch, 1962; Gambetta, 1988; Luhmann, 1979). Trust can be considered a variate within society; "sometimes there is more and sometimes there is less" (Gellner, 1988, p. 142). Decisions to trust are laden with risk (Luhmann, 1988; Sztompka, 1999). There is always the risk that something we trust and/or "those whom we trust will not act as expected" (Govier, 1992, p. 17).

Recent research has provided valuable insights relevant to schools in terms of the dynamics of trust and why people trust (Kramer & Tyler, 1996). Studies of interpersonal trust (Gambetta, 1988; Lewis & Weigert, 1985; Tarter, Sabo, & Hoy, 1995; Tschannen-Moran & Hoy, 1998; Kramer & Tyler, 1996; Zucker, 1986) have appeared with regularity in related discipline-based and professional literatures. In contrast, few studies focus directly on institutional trust. The minimal research on institutional trust of public institutions, including schools, suggests "a serious analytical shortfall" in terms of "our understanding of the bases for institutional trustworthiness, and about the conditions that would reduce serious deficits in public trust and confidence—a situation akin to betrayal in individual relations" (LaPorte & Metlay, 1996, p. 342).

In schools, institutional trust may transcend the concrete exchange experience, and be diffused in a broader set of relationships (Kramer & Tyler, 1996; Lane & Bachmann, 1998). Schools, as social institutions within social systems, may be viewed as administrative organizations that are "not reducible to features of individual actors" (Whitley, 1987, p. 133). Trust in schools as structural arrangements and trust in their institutionalized practices and procedures may be expected to vary over time.

Schools, as objects of trust, may depend on the establishment and maintenance of confidence that good results will be achieved if practices and procedures are

followed (Sztompka, 1999). These institutions may be strengthened where referent publics' bets on efficient and competent operations pay off; for example, where reliable professional services are provided, where schools operate in an orderly manner, where resources are appropriately managed, and where agents show their trustworthiness by placing public interests before their own (Barber, 1983; Erikson, 1987). Trust in schools may be reinforced, moreover, where the legitimacy of the authority of schools is strengthened by broad-based citizen confidence in those responsible for exercising authority, and in the institutions themselves. Trust may take the form of reliance on these social institutions as entities capable of delivering on citizen expectations (Barber, 1983; Giffin, 1967).

Alternative Lenses and Constructs

Lewicki and Bunker (1996) suggested that existing perspectives on trust may be aggregated in three categories: views of personality theorists who have focused on individual personality differences in the readiness to trust, and on the specific developmental and social contextual factors that shape this readiness; views of social psychologists who have focused on interpersonal transactions between individuals that create or destroy trust at the interpersonal or group level; and views of sociologists and economists, who have focused on trust as an institutional phenomenon (here, trust is conceptualized as both a phenomenon within and between organizations, and as the trust individuals place in those institutions).

A review of research on trust in schools and other institutions shows decided concern for interpersonal trust and its importance in terms of maintaining individual and organizational effectiveness (Bruhn, 2001; Butler, 1991; Byrk & Schneider, 2002; McAllister, 1995; Whitener, Brodt, Korsgaard, & Werner, 1998). Employee trust in subordinate/ managerial relationships has also received considerable attention (Butler & Cantrell, 1984; Keefe, 1992; Lagace, 1991; McCauly & Kuhnert, 1992). Research on trust applicable to schools (Tschannen-Moran & Hoy, 2000), more specifically, appears to be skewed in favor of examinations of relational aspects of the phenomenon. Numerous studies of schools focus on trust and related effects on operational effectiveness, productivity, student achievement, overall smooth organizational functioning, positive climate, productive communication and participative decision-making (Bandura, 1993, 1997; Henkin & Dee, 2001; Tschannen-Moran & Hoy, 2000).

Perspectives shift when schools themselves become the focal concern. The study of trust between an individual and an organization suggests alternative views of trust at several levels of analysis (Barber, 1983; Bryan, 1995; Mishra, 1992). Reputation as a measure of trust, for example, is commonly cited as one basis for determining the trustworthiness of a target (Sztompka, 1999), and may be the basis for decisions about whether or not to trust someone or something (Good, 1988; Zucker, 1986). "Reputation is a symbolic representation of past exchange history" (Zucker, 1986, p. 54); a "record of past deeds" (Sztompka, 1999, p.71); "an acquired capital asset" (Dasgupta, 1988, p. 50); and a marker of "a certain unity of conduct over time" (Sztompka, 1999, p. 72). Constancy of action over time suggests the capacity and willingness to continue to perform in the same way over time (La Porte & Metlay, 1996).

Trust may be considered as a moderating variable among a number of factors that impact individual-institutional orientations. Credibility and benevolence, and a record of fair treatment of clients may suggest, for example, that the provider (here, service provider) can be relied upon (credibility) and has the clients' best interest in mind (benevolence) (Mayer, Davis & Schoorman, 1995; Wilson, 1995). Nonetheless, organizational trust, like forms of relational trust, involve risk on the part of the trusting party (Lane & Bachmann, 1998); risk being one of " the few characteristics common to all trust situations" (Mayer, Davis, & Schoorman, 1995, p. 712).

Multiple definitions of trust may be useful in research on institutional dimensions (Bhattacherjee, 2002). Similarities across definitions suggest certain conceptual commonalities. The construct is frequently linked to positive expectations in terms of the behavior of person(s) and/or the functioning of an entity (like a school). Trust dimensions found in definitions often reflect core beliefs that define a perspective; for example, promise fulfillment, openness, and fairness, among others (Kramer & Tyler, 1996; Bhattacherjee, 2002). Trust as an expectation suggests, additionally, a property of the phenomenon as a future state (Barber, 1983: Zucker, 1986) where there is predictability of performance in terms of the object of trust (Giffin, 1967). The concept may be applied to a school that citizens deem to be trustworthy.

The idea of institutional trust continues to seek definition through lenses that may include a range of contextually mediated characteristics or attributes. For schools, we may include characteristics such as goodwill, performance consistency, reliability, and predictability, among others (Barber, 1983; Giffin, 1967). Institutional trust may find some level of acceptable definition as a multidimensional construct with multiple properties that are differentiated within perceptions of operational functioning (Mayer, Davis, & Schoorman, 1995; McKnight & Chervany, 1996), in contrast with other forms of trust that may seek definition in behavioral manifestations in interpersonal interactions and relationships.

Trust, conceptualized as confidence, is found in many definitions (Barber, 1983; Lewicki, Stevenson, & Bunker, 1997; Luhmann, 1979; Mayer, Davis, & Schoorman, 1995; Zucker, 1986). Trust in schools may emerge, independent of suggested characteristics, via legalistic mechanisms and institutionally associated forms such as formal social structures that may influence views of organizational trustworthiness (Lewis & Weigert, 1985; Sitkin & Roth, 1993; Zucker, 1986). Lewis and Weigert (1985) suggested that with increasing population growth and structural

differentiation, an increasing number of social relationships may come to be based on institutional rather than interpersonal trust. Modern trust may be more closely linked to a person's sense of how institutions, like schools, operate, rather than to individual attitudes about, and confidence in specific institutions populated by known individuals (Sellerberg, 1982). Citizen trust in schools can take the form of expressions of confidence in operations in accordance with law and bureaucracy (Govier, 1997). Trust may be extended to institutions where there may be limited information about individuals and authorities; a necessity in societies where the pressures and realities of modern life provide insufficient time for building up the number of personal trusts required in alternative models (Lewis & Weigert, 1985).

Role-based trust (Barber, 1983; Dawes, 1994) may serve as a proxy for personalized knowledge of institutions and their technologies. In role-based trust, there is confidence and a presumption of trust in terms of expectations of technically competent role performance, and confidence and trust that role incumbents have the intent to fulfill their obligations. Trust is presumed without the benefits of personalized knowledge or prior interaction. In instances of presumptive trust, it is often trust in the system of expertise that is trusted rather than the role incumbent her-/himself. The system of expertise is presumed to produce and maintain role-appropriate behavior (Dawes, 1994). " We trust engineers because we trust engineering and believe that engineers are trained to apply valid principles of engineering. Moreover, we have evidence every day that these principles are valid when we observe airplanes flying" (p.24). Where role-based trust is extended to schools, there may be significant variation in terms of a presumption of trust in terms of intentions and capabilities. The extent to which there may be trust in the system of expertise may remain indeterminate. Here, role-based trust may be a fragile proxy for personalized knowledge.

Personal trust may cede to a kind of trust in the ability of institutions like schools to maintain conditions of performance (Govier, 1997); that is, to forms of impersonal trust that are embedded in the operations of institutions rather than in the structures of interpersonal relationships (Shapiro, 1987). Trust may be activated, for example, by the appearance that everything is in proper order (Luhmann, 1979).

Schools as institutions may be designed to enhance perceptions of trustworthiness that makes trusting successful (Hardin, 1996). Whitener and her associates (1998) suggested specific elements or structures that strengthen the likelihood of institutions being perceived as trustworthy; for example, high levels of decentralization, low formalization, the design of human resource policies and procedures following principles of procedural justice, and development of an organizational culture characterized by inclusiveness and open communication; characteristics that may be found in effective self-managed schools. Trust may emerge, furthermore, via the confidence people have in a situation because of guarantees, safety nets and certain familiar structures (Lewis & Weigert, 1985; Shapiro, 1987). Schools with guarantees suggested in structure-based trust models may mitigate perceptions of risk linked to notions of fairness.

Institution-based trust may be supported by citizen views that situations are bounded by organizational safeguards. Such safeguards may strengthen confidence that individuals in the organization are trustworthy. For example, the benevolence of schools as employers may be deemed more credible where there are provisions for sanctioning abusive administrative practices. Additionally, beliefs in the likelihood that situations will remain normal and people will act normally, consistent with research on cognitive consistency (Abelson, Aaronson, McGuires, Newcomb, Rosenberg, & Tannenbaum, 1968), support the assertion that institutionally-related trusting beliefs may remain stable over long periods of time.

Drawing on the work of Shapiro (1987) and Zucker (1986), McKnight, Cummings, and Chervany (1998) suggested that institution-based trust means that one believes the necessary impersonal structures are in place to enable one to act in anticipation of a successful future endeavor (p. 477). Two related conditions for institution-based trust are discussed in the literature; specifically, situational normality and structural assurances. McKnight and associates (1998) asserted that situational normality exists where institutional settings appear consistent with institutional purposes. Trust may break down, in contrast, when individuals face inexplicable, abnormal situations (Garfinkel, 1963). For schools, citizen trust may be strengthened where structural assurances and safeguards reinforce confidence that both the institution and institutional authorities are trustworthy (McKnight et al., 1998).

The ability of school authorities to develop trust relationships with citizens is constrained where opportunities for interaction are necessarily limited (Hart, 1989); the case in many communities. Viewed from an institutional perspective, trust in schools and their authorities may emerge from judgments about executive decisions, and related impacts on the institution and community. Institutional trust may be associated, in this instance, with citizen perceptions of institutional efficiency and fairness, rather than with personal characteristics and behaviors of executives (McCauley & Kuhnert, 1992). Citizens may consider the efficiency and fairness of school policies and procedures, for example, as they form beliefs in terms of institutional trustworthiness (McCauley & Kuhnert, 1992; Fox, 1974). The importance of institutional provisions and conditions that reinforce individual confidence in entities, including schools, is suggested (Kramer & Tyler, 1996; Sztompka, 1999).

Individual judgments about institutional trustworthiness may be significantly affected by violated expectancies (LaPorte & Metlay, 1996; Sztompka, 1999). Where there are violations of public trust by key organizational personnel, there may be decreases in trust extended beyond the violators

to the broader institution. Reactions to violated trust may be particularly acute among referents who had placed high levels of trust in the violators. When institutional trust is defined on the basis of behaviors of highly visible violators of trust, moreover, the violations may be generalized to redefine the trustworthiness of associated institutions and systems. Related research on distrust and suspicion (Nye, Zelikow, & King, 1997) that focuses on perceived failures of public agencies to deal effectively with, or solve some of the "wicked problems" of society suggests, similarly, a tendency among referents to redefine failures as generalized distrust. Unmet expectancies may significantly impact perceptions of institutional trustworthiness.

Inferences of this line of research for trust of public schools may be suggested. Where school leaders are focal points for gauging beliefs about institutional trustworthiness, for example, perceptions of misplaced trust and/or trust violation linked to these leaders may rapidly extend to the institution(s) themselves. The pace of erosion of trust may be most acute among referents who place the highest levels of trust in both the school leadership and the institution itself. Similarly, continuous failures of schools to respond to referent expectancies may function to redefine, in part, the extent to which schools are perceived as trustworthy community assets.

Elements in an Institutional Trust Framework

A theoretically grounded, initial framework for institutional trust may suggest factors important in reducing citizen uncertainty and strengthening perceptions of schools as trustworthy organizations. Schools perceived as trustworthy, will be in a better position to attract higher levels of community support and sustained public investment.

People make trusting decisions many times each day (Luhmann, 1979). Decisions to trust are usually based on beliefs or confidence in individuals' or organizations' (here, a school) good intentions toward us (Yamamoto, 1990). When we are disappointed in the object of trust, we tend to trust that thing less (or not trust it at all). Disappointment on the part of the trustee is not within our realm of responsibility. It is the responsibility of the trustee (Hertzberg, 1988), since "trust can only concern that which one can rightly demand of another" (p. 319).

In the public trust equation, institutional constancy appears to be a major issue. An institution shows constancy when it consistently achieves goals that it agreed to pursue in the past. Educational institutions may display constancy in commitments to academic excellence, in evidence where intellectually prepared, skilled graduates are produced year after year (LaPorte & Metlay, 1996). Where there is demonstrated, consistent, high level performance that is synchronous with public expectations, there are assurances for citizens that institutions are trustworthy. They may infer that an institution will be faithful to commitments, consistent in performance, and exhibit characteristics and qualities that demonstrate willingness and capability to do so into the indefinite future (LaPorte & Metlay, 1996). Institutional constancy may be a significant element in trust that functions, theoretically, as a means that citizens may employ to simplify social complexity in technically complex social systems (Luhmann, 1979). Individuals have a need to trust institutionalized elements in a complex system in order to cope with the complexities of individual experience within the system (Barber, 1983; Luhmann, 1979; Shapiro, 1987).

The development of trust (or distrust) in organizations tends to be cumulative over time. "All organizations have some pockets of distrust at some time in their life cycles" (Bruhn, 2001), and that distrust may be evident as a residual after the source of distrust is no longer present (Morgan, 1997). We reason, similarly, that perceptions of trustworthiness are strengthened over the longer term where constancy of performance consistent with public expectations is the norm. Constancy may be reinforced by dimensions of trust (ability, benevolence, and integrity) suggested in a typology that appears relevant to schools (Mayer, Davis & Schoorman, 1995). These dimensions, while conceptually distinct, "represent a comprehensive yet parsimonious dimension space for trust formation" (Bhattacherjee, 2002, p. 217). A school that predictably demonstrates minimal variation over time in terms of willingness and ability to serve the community may be perceived as worthy of trust.

Perceptions of trustworthiness related to the ability dimension of trust may emerge via signals of competence (here, the ability to do what has to be done) and by demonstrating a willingness to perform in expected ways. A school can show, for example, that it can consistently enact effective processes to deliver services through systems populated with professionals who have the knowledge and skills required to perform at high levels of proficiency within their domains of expertise.

The benevolence dimension (here, showing a concern for the welfare and interests of others) may appear in school organizations as a demonstrated, consistent willingness to serve others. Benevolence may be shown, for example, through expressions of interest in, and empathy toward community challenges and needs, by being active in efforts to address larger community problems and concerns, and by introducing a level of altruism into the trust equation through demonstration of consistent inclination to function in good faith and contribute when unanticipated community needs may arise. Such actions counter perceptions of self-interest and opportunism, and strengthen beliefs in institutional benevolence.

The integrity domain involves a trustor's confidence, and perceptions that the trustee will abide, in the case of schools, by principles and rules that the trustor deems to be both fair and reasonable. Integrity in for-profit settings has been equated with such corporate attributes as openness, honesty, forthrightness, and a willingness to voluntarily disclose information; even when disclosure may result in economic loss (Paine, 1994). While parallel research focused

on traditional social institutions like schools is not available, we suggest that integrity as a antecedent of trust may be strengthened where organizational actions confirm values and principles that reinforce organizational credibility and promote perceptions of trustworthiness (Peters, Covello, & McCallam, 1977).

Finally, we suggest the need to include variables of risk and individual trustor propensity to trust in an institutional trust framework (Gambetta, 1988; Mayer, Davis & Schoorman, 1995; Tyler & Kramer, 1996). Citizens take risks and become vulnerable when they trust schools. They show a willingness to be vulnerable, when they risk trust in schools as agencies. Simply put, they expect the trustee to perform. Actions (performance) of the trustee occurs independent of any capacity of the trustor to monitor those actions. Implicit in the trust-risk concept, then, is trust based on trustor assessment of the subjective probability that the trustee (school) will perform expected actions, and the that the actions will be beneficial to the trustor. Expectancy theories of trust (Dasgupta, 1988; Gambetta, 1988; Kramer & Tyler, 1996) suggest that there are incentives to be trustworthy that may (or may not) be in place at the time of the need for fulfillment of trustor expectation. The reciprocity factor in citizen-school relationships should not be neglected, as we progress toward better understandings of institutional trust.

Directions for Research

Research on institutional trust deals with issues of organizations as sources of trust, and with individual belief and confidence in the abstract system (Lane, 1998). To this point, we have suggested how trust may be embedded in formal structures and operations of schools.

Consideration of organizational effects in personality level research on trust would enhance our understanding of trustors' characteristics that may function as filters through which levels of institutional trustworthiness are determined. Related research on the nature and operation of attitudes (Ajzen, 2001), moreover, could provide a theoretically grounded foundation for understanding how citizen trust beliefs are formed. Such research may provide insights, for example, into how information about schools, consistent with the theory of reasoned action (Ajzen, 1985, 2001; Fishbein & Ajzen, 1975), may impact the development of beliefs that affect attitudes (favorable or unfavorable) toward the school; attitudes that may be manifest in behaviors related to public issues significant for schools. The theory, applied in our case to trust of schools, may suggest patterns where beliefs/attitudes (trusting beliefs) lead to intentions (trusting intentions) that, subsequently, become manifest in behaviors (trust of the school). Further research on institutional trust, considerate of this line of reasoning, may yield findings with significant implications in terms of potential support for community schools deemed more or less trustworthy (Ajzen, 2001; Ajzen & Driver, 1992; East, 1993; Hofstede, 1980).

Trust in schools and other social organizations of modern society may depend, in part, on a range of minimally considered variables associated with impersonal dimensions of the phenomenon. It is suggested, here, that high levels of citizen trust in schools as institutions may generate more favorable attitudes, and influence individual decisions related to schools that are conditioned by anticipation of positive outcomes. Schools perceived as trustworthy may serve as positive forces in larger social systems where incomprehensible political and economic institutions of government appear, increasingly, to be contributing to citizen alienation that may be manifest in crises of confidence.

We acknowledge the need to consider the potential effects of intervening variables that may function as antecedent conditions of impersonal forms of trust; for example, diversity of community cultural backgrounds that may impact formation of trust relationships, and individual value orientations that may influence what information may be processed and found credible (Gudykunst, 1997; Hofstede, 1980; Nicolau & Ramos, 1990). There may be social and culturally-bound dispositions not to trust schools, for example, and additional mitigating risk factors associated with citizens' lack of first-hand knowledge about schools (Triandis, 1989; Yamagishi & Yamagishi, 1994).

This exploration of institutional dimensions of trust, trust extended beyond personal experience to impersonal structures in the context of modern society (Luhmann, 1979; Giddens, 1993), suggests the importance of theoretical perspectives and related constructs as points of departure for research. We acknowledge the utility of multiple disciplinary lenses, and recognize that this line of research may be further complicated by variation in conceptualizations of the trust construct (Mayer, Davis, & Schoorman, 1995; Rousseau, Sitkin, Burt, & Camerer, 1998) and by inconsistencies in asserted associations between trust and its antecedents and consequents (Rousseau, et al., 1998; Kramer & Tyler, 1996). Future inquiry is challenged, moreover, to generate study designs that include community-specific variables important in the school trust equation.

Schools may be viewed as organizations endowed with trust. Trust in these institutions may be expected to vary over time (Sztompka, 1999). The importance of research on institutional trust for resource-dependent schools becomes all the more apparent when we consider the potential impact of diminished trust or distrust in schools and their agents.

References

Abelson, R., Aaronson, E., McGuires, W., Newcomb, T., Rosenberg, M., & Tannenbaum, P. (Eds.). (1968). *Theories of cognitive consistency: A sourcebook.* Chicago: Rand-McNally.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior*, (pp.11-39). New York: Springer Verlag.

Ajzen, I. (2001). Nature and operation of attitudes. *Annual*

Review of Psychology, 52, 27-58.

Ajzen, I. & Driver, B. (1992). Application of the theory of planned behavior to leisure choice. *Journal of Leisure Research, 24*, 207-224.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Barber, B. (1983). *The logic and limits of trust.* New Brunswick, NJ: Rutgers University Press.

Bhattacherjee, A. (2002). Individual trust in on-line firms. *Journal of Management Information Systems, 19*(1), 211-242.

Beeson, G., & Matthews, R. (1993). *Collaborative decision making between new principals and teachers: Policy and practice.* ED 361837.

Bok, S. (1978). *Lying: Moral choice in public and private life.* New York: Pantheon Books.

Bruhn, J.(2001). *Trust and the health of organizations.* New York: Kluwer.

Bryan, G. (1995). *The dynamics of individual-to-organizational trust relationships in a higher education setting.* Unpublished dissertation, Cleveland State University, Cleveland, OH.

Bryk, A. & Schneider, B. (2002). *Trust in schools: A core resource for improvement.* New York: Russell Sage.

Butler, J. (1991). Towards understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management, 17*, 643-663.

Butler, J., & Cantrell, R. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports, 55*, 19-28.

Coleman, J. (1990). *Foundations of social theory.* Cambridge, MA: Belknap Press of Harvard University Press.

Creed, W. & Miles, R. (1996). Trust in organizations: A conceptual framework linking organizational forms, managerial philosophies, and the opportunity costs of controls. In R. Kramer & T. Tyler (Eds.), *Trust in organizations* (pp. 16-38). Thousand Oaks, CA: Sage.

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (p. 49-72). New York: Basil-Blackwell.

Dawes, R. (1994). *House of cards.* New York: Free Press.

Dee, J. & Henkin, A. (2001). *Smart school teams: Strengthening skills for collaboration.* Lanham, MD: University Press of America.

Deutsch, M. (1962). Cooperation and trust. In M. Jones (Ed.), *Nebraska symposium on motivation* (pp. 275-317). Lincoln, NE: University of Nebraska Press.

East, R. (1993). Investment decisions and the theory of planned behavior. *Journal of Economic Psychology, 14*, 337-375.

Elmore, R. (1993). *Education, decentralization, and school improvement: Can we fulfill the promise?* San Francisco: Jossey-Bass.

Erikson, F. (1987). Transformation and school success: The politics and culture of educational achievement. *Education and Anthropology Quarterly, 18*, 16-24.

Fishbein M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior.* Reading, MA: Addison-Wesley.

Fox, A. (1974). *Beyond contract: Work, power, and trust relations.* London: Faber & Faber.

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust : Making and breaking cooperative relations* (pp. 213-238). New York: Basil-Blackwell.

Garfinkel, H. (1963). A conception of, and experiments with trust as a condition of stable, concerted actions. In O. Harvey (Ed.), *Motivation and social interaction* (pp.187-238). New York: Ronald Press.

Gellner, E. (1988). Trust, cohesion and the social order. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*, (pp. 142-157). New York: Basil-Blackwell.

Giddens, A. (1993). *Modernity and self identity.* Cambridge: Polity Press.

Giffin, K. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin, 68*(2), 104-120.

Good, D, (1988). Individuals, interpersonal relations, and trust. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*, (pp. 31-48). New York: Basil-Blackwell.

Govier, T. (1992). Trust, distrust and feminist theory. *Hypatia, 7*(1), 16-33.

Govier, T. (1997). *Social trust.* London: McGill-Queen's University Press.

Gudykunst, W. (1997). Cultural variability in communication. *Communication Research, 24*, 327-348.

Hardin, R. (1996). Trustworthiness. *Ethics, 107*, 26-42.

Hart, K. (1989). A requisite for employee trust: Leadership. *Psychology, 25*(2), 1-7.

Henkin, A., & Dee, J. (2001). The power of trust: Teams and collective action in self-managed schools. *Journal of School Leadership, 11*, 49-58.

Hertzberg, L (1988). On the attitude of trust. *Inquiry, 31*, 307-322.

Hofstede, G. (1980). *Culture's consequences.* Beverly Hills, CA: Sage.

Jones, W., & George, J. (1998). The experience and evolution of trust. *Academy of Management Review, 23*, 531-546.

Keefe, J. (1992). *School restructuring.* Reston, VA: NASSP.

Kramer, D., Brewer, M., & Hanna, B. (1996). Collective trust and collective action. In R. Kramer & T. Tyler (Eds.), *Trust in organizations* (pp. 357-389). Thousand Oaks, CA: Sage.

Kramer, R., & Tyler, T. (Eds.) (1996). *Trust in organizations.* Thousand Oaks, CA: Sage.

Kramer, R., & Tyler, T. (1996). Whither trust. In R. Kramer & T. Tyler (Eds.), *Trust in organizations* (pp. 1-15).

Thousand Oaks, CA: Sage.

Lagace, R. (1991). An exploratory study of reciprocal trust between sales managers and salespersons. *Journal of Personal Selling and Sales Management, 11*, 49-58.

Lane, C. (1998). Introduction: Theories and issues in the study of trust. In C. Lane & R. Bachmann (Eds.), *Trust within and between organizations* (pp. 1-30). New York: Oxford University Press.

Lane, C. & Bachmann, R. (Eds.). (1998). *Trust within and between organizations*. New York: Oxford University Press.

LaPorte, T. & Metlay, D. (1996). Hazards and institutional trustworthiness: Facing a deficit of trust. *Public Administration Review, 56*(4), 341-348.

Lewicki, R., & Bunker, B. (1996). Developing and maintaining trust in working relationships. In R. Kramer & T. Tyler (Eds.), *Trust in organizations* (pp.114-139). Thousand Oaks, CA: Sage.

Lewicki, R, Stevenson, M. & Bunker, B. (1997). *The three components of interpersonal trust*. Paper, Conflict Management Division, Academy of Management National Meeting.

Lewis, J., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63*(4), 967-985.

Luhmann, N. (1979). *Trust and power*. New York: John Wiley.

Luhmann, N. (1988). Familiarity, confidence, trust. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations*, (pp. 94- 107). New York: Basil Blackwell.

Mathews, D. (1996). *Is there a public for public schools?* Dayton, OH: Kettering.

Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.

McAllister, D. (1995). Affect and cognition based trust as foundations for cooperation in organizations. *Academy of Management Journal, 38*, 24-59.

McCauley, D., & Kuhnert, K. (1992, Summer). A theoretical review and empirical investigation of employee trust in management. *Public Administration Quarterly*, 265-284.

McKnight, D, & Chervany, N. (1996). The meanings of trust. Working Paper. (www.misrc.umn.edu/wpaper/wp96-04.htm).

McKnight, D., Cummings, L., & Chervany, N. (1998). Initial trust formation in new organizational relationships. *Academy of Management, 23*(3), 473-490.

Mishra, A. (1992). *Organizational responses to crisis: The role of mutual trust and top management teams*. Unpublished dissertation, University of Michigan, Ann Arbor.

Morgan, G. (1997). *Images of organization*. Thousand Oaks, CA: Sage.

Nicolau, S. & Ramos, C. (1990). *Together is better*. Washington, DC: Hispanic Policy Development Project.

NORC. (November, 1993). Public opinion and demographic report: Confidence in institutions. *American Enterprise, 4*, 94-95.

Nye, J., Zelikow, P. & King, D. (1997). *Why people don't trust government*. Cambridge, MA: Harvard University Press.

Paine, L. (1994). Managing for organizational integrity. *Harvard Business Review, 72*(2), 106-113.

Peters, R., Covello, V., & McCallum, D. (1997). A study of trust and credibility factors. *Risk Analysis, 17*(1), 43-54.

Pounder, D. (1998). *Restructuring schools for collaboration*. Albany, NY: SUNY Press.

Putnam, R. (1993). *Making democracy work*. Princeton, NJ: Princeton University Press.

Purdy, J. (2003). Suspicious minds. *Atlantic Monthly, Jan/Feb*, 82-86.

Reyes, R. (Ed.). (1990). *Teachers and their workplace*. Newbury Park, CA: Sage.

Rittel, H. & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155-169.

Rossow, J., & Zager, R. (Eds.). (1989). *Allies in educational reform*. San Francisco: Jossey-Bass.

Rotter, J. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443-452.

Rousseau, D., Sitkin, S., Burt, R., & Camerer, L. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Sellerberg, A. (1982). On modern confidence. *Acta Sociologica, 25*, 39-48.

Shapiro, P. (1987). The social control of impersonal trust. *American Journal of Sociology, 93*, 623-658.

Simon, H. (1991). Organizations and markets. *Journal of Economic Perspectives, 5*, 34-38.

Sitkin, S. & Roth, N. (1993). Explaining the limited effectiveness of legalistic remedies for trust/distrust. *Organization Science, 4*(3), 367-392.

Sztompka, P. (1999). *Trust*. Cambridge, UK: Cambridge University Press.

Tarter, C., Sabo, D., & Hoy, W. (1995). Middle school climate, faculty trust and effectiveness: A path analysis. *Journal of Research and Development in Education, 29*, 41-49.

Triandis, H. (1989). Cross-cultural studies of individualism-collectivism. In J. Berman (Ed.), *Symposium on motivation: Cross cultural perspectives*. Lincoln, NE: University of Nebraska Press.

Tschannen-Moran, D. & Hoy, W. (1998). A conceptual and empirical analysis of trust in schools. *Journal of Educational Administration, 36*, 334-352.

Tschannen-Moran, M., & Hoy, W. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research, 70*(4), 547-593.

Tyler, T. (2000). Why do people rely on others? Social identity and social aspects of trust. In K. Cook (Ed.), *Trust in society* (pp. 285-306). New York: Russell Sage.

Watson Wyatt Worldwide. Work USA 2002- Weathering

the storm: A study of employee attitudes and opinions. Washington, DC: Watson Wyatt.

Whitener, E., Brodt, S., Korsgaard, M., & Werner, J. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review*, 513-530.

Whitley, R. (1987). Taking firms seriously as economic actors: Towards a sociology of firm behavior. *Organizational Studies, 8*(2), 125-147.

Wilson, D. (1995). An integrated model of buyer-seller relationships. *Journal of the Academy of Marketing Science, 23*(4), 335-345.

Yamagishi, T. & Yamagishi, M. (1994). Cross cultural differences in risk perception, but cross-cultural similarities in attitudes. *Management Sciences, 44*, 1205-1217.

Yamamoto, Y. (1990). A morality based on trust. *Philosophy East and West, 40*(4), 451-469.

Zucker, L. (1986). The production of trust: Institutional sources of economic structure. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (pp. 55-111). Greenwich, CT: JAI Press.

Alan Henkin is Professor of Educational Policy and Leadership Studies at The University of Iowa. His research interests include organizational behavior and analysis, policy development and implementation, and management theory and practice. Recently published books and articles focus on interpersonal and systems trust, commitment, empowerment, and collective action.

Carole A. Singleton is Professor of Social Work at the University of Northern Iowa. Her teaching, research and publications focus on administration in social agencies, social communications skills, collaboration, conflict mediation, and trust building.

Fred Holman is Dean of Continuing and Adult Education at St. Thomas University, Miami, Florida. His research interests include administrative leadership in schools and colleges, executive skills development, and school higher education collaboration.

Jay Dee is Assistant Professor of Leadership in Education in the Graduate College of Education at the University of Massachusetts-Boston. His research and teaching focus on organizational theory, governance, faculty roles and reward systems in complex organizations, and trust and teamwork in schools.

A Lesson Study Model for Preservice Teacher Education

Tracy C. Rock
University of North Carolina

This paper presents a lesson study model for preservice teacher education that was adapted from the Japanese lesson study process described by Stigler and Hiebert (1999) in The Teaching Gap. In addition, data from multiple case studies describes how preservice teachers perceived the model and its effects on their understandings and dispositions toward their own professional growth and development. The prospective teachers in this study found the lesson study process to be a rewarding professional development experience. They perceived the lesson study experience to be a unique opportunity to observe, analyze, and understand how changes made to an initial lesson plan actually impacted the effectiveness of the lesson. The findings of this study also indicate that when preservice teachers work collaboratively with peers through the lesson study process they gain knowledge and skills of collaboration and experience its value in enhancing professional understandings and practices.

Introduction

Preservice teachers should understand early in their professional coursework that learning to teach is a career-long endeavor rather than something they will accomplish in a preservice teacher education program. Therefore, professional education courses should provide opportunities for preservice teachers to develop skills of inquiry, reflection, problem solving, and collaboration for the purpose of on-going professional growth (Darling-Hammond, 1998; Holmes Group, 1990). Schon (1983) emphasized the processes of reflection in action and reflection on action as necessary elements for the continuous improvement of teaching. Darling-Hammond (2003) states, "Good settings for teacher learning - in both colleges of education and schools- provide lots of opportunities for research and inquiry, for trying and testing, for talking about and evaluating the results of learning and teaching" (p. 278).

This paper presents a lesson study model for preservice teacher education that was developed as a semester-long field experience assignment for an elementary social studies methods course. The model was adapted from the Japanese lesson study process described by Stigler and Hiebert (1999) in *The Teaching Gap.* Lesson study was selected and modified as a structure for the field experience because it provides a framework that encourages the development of skills and habits of mind necessary for effective teaching and professional growth. In addition, data from multiple case studies are shared to describe how preservice teachers perceived the model and its effects on their understandings and dispositions toward their own professional growth and development.

Background of Lesson Study

In the *The Teaching Gap,* Stigler and Hiebert (1999) draw on the conclusions of the Third International Math and Science Study (National Center for Education Statistics, 1996), an innovative study that examines teaching practices in several cultures, to set forth a plan for change in America's schools. Their work centers on restructuring schools to become places where teachers can engage in career-long learning and classrooms can become laboratories for developing new, teacher-centered ideas for improving student achievement. They advocate that by providing time for lesson study, a professional development process that allows teachers to target aspects of their instruction and collaboratively study how to improve that instruction, teachers will change the way our students learn.

Despite years of reform, research suggests there has been little change in classroom teaching in the United States (National Center for Education Statistics, 1996). In contrast, teaching practices in Japan appear to have dramatically changed over the past 50 years (Lewis & Tsuchida, 1997). Typically, U.S. reform efforts seek major changes over relatively short time periods; whereas, Japanese educators have instituted a system that brings gradual, incremental improvements to teaching over time.

In Japan the primary responsibility for the improvement of classroom practice is given to the teachers (Lewis & Tsuchida, 1998). Kounaikenshuu is the word used to describe the continuous process of professional development occurring within Japanese schools. Lesson study is one of the most common components of Kounaikenshuu (Lewis, 2000). Lesson study (jugyou kenkyuu) involves groups of

teachers meeting regularly over a period of time (ranging from several months to a year) to construct, implement, test, and improve one or several research lessons (Stigler & Hiebert, 1999). This activity is reported to be the "linchpin of the improvement process"(Stigler & Hiebert, 1999, p. 111) in the Japanese educational system. The major idea is that if you want to improve teaching then the most appropriate place to do so is within the context of a classroom lesson. The steps involved in the lesson study process utilized in Japanese schools include (a) defining and researching a problem of practice, (b) planning, teaching, and observing the lesson, (c) evaluating the lesson, (d) revising the lesson, (e) teaching and observing the revised lesson, (f) evaluating and reflecting again, and (g) sharing the results.

Research Questions

The purpose of this initial line of inquiry was to describe how preservice teachers perceived the lesson study process as a professional development tool. The research questions for this study were (a) How do preservice teachers perceive the lesson study model? and (b)What are the benefits and constraints of engaging in lesson study collaboratively as compared to working individually?

Method

Context

All the professional courses within the University of North Carolina at Charlotte's elementary teacher education program require a field experience. These field experiences are carried out in multiple field sites assigned to the preservice teachers as they move through the program. As seniors the preservice teachers are assigned to a yearlong internship placement in an elementary school where they conduct all first-semester professional course field experiences and in the second-semester carry out their student teaching requirements. Many preservice teachers live in surrounding counties and request field sites and yearlong internship placements near their homes, therefore, the preservice teachers are often working in a variety of locations.

A Collaborative Lesson Study Model for Preservice Teacher Education

In this section, a modified lesson study model for preservice teacher education is presented. The intent is to engage preservice teachers to collaboratively planning and examining lessons, so they can experience the value of such a process to their own development as a teacher and in improving teaching and learning in their classroom. The semester-long field experience is divided into six phases:

Phase I. Following an overview of the lesson study process, preservice teachers form their own partnerships within the methods class based on grade-level field assignments and schedule compatibility. Prior to this experience these preservice teachers have little to no experience teaching social studies in the elementary school classroom; therefore, unlike experienced teachers, they are unable to formulate questions revolving around areas of subject matter difficulty or common social science misconceptions of elementary-aged children. They are initially concerned with learning the content that is to be taught within the curriculum and experimenting with effective ways to teach it. During this preliminary phase they must grapple with the questions: What is it that we need to teach? How will we teach the material effectively? The collaborative partners along with their cooperating teachers engage in discussing, negotiating, assessing possibilities, and examining constraints as they identify the focus of the lesson study.

Once the focus of the lesson study has been identified the preservice teachers engage in lesson study research. Each partner gathers information from three sources. First, the preservice teachers must interview a diverse group of students from their individual field placements to better understand the students' thinking and prior knowledge of the concepts to be taught during the lesson study. Next, they gather at least three articles from the professional literature related to the content and/or methodology for the lesson study and they read, summarize, and reflect on implications for practice. Lastly, they observe at least two social studies lessons taught by their cooperating teacher and prepare a structured written reflection of the two lessons. Once the lesson study research is complete they are ready to move on to Phase II.

Phase II. The collaborative partners meet to share knowledge gained from students, professional literature, and co-op teachers. After discussing what has been learned the partners begin to plan and write an initial formal lesson plan to be taught. At this point the preservice teacher partners consult with the professor about their initial lesson plan. When the initial lesson plan is finalized then they are ready to move to Phase III.

Phase III. During phase III, each participant has different roles and responsibilities in the lesson study process. Partner A teaches the initial lesson in assigned placement and has it videotaped. After the lesson is taught, Partner A reviews the video and prepares a structured written narrative reflection from the videotape. Also, a peer coaching feedback form is completed. Partner B comes to observe the initial lesson taught and assists with the videotaping. Following the lesson, Partner B completes a peer coaching form also. The cooperating teacher observes the lesson and provides written feedback in the form of strengths and suggestions for growth. The cooperating teacher is also encouraged to provide verbal encouragement and suggestions.

Phase IV. The collaborative partners (Partner A & B) meet to share and discuss what was learned from the implementation of the initial lesson plan. During this phase the preservice teachers are required to reflect, rethink, and replan. Based on their new understandings of the lesson, the preservice teachers create a revised lesson plan.

Phase V. The next step is to repeat Phase III with the revised lesson plan. During Phase V, Partner A and B switch roles and responsibilities. Partner B will now teach the revised lesson in their field placement with their assigned cooperating teacher providing written and verbal feedback. Partner B completes a peer observation form and writes a narrative reflection of the revised lesson. Partner A observes the revised lesson being taught, assists with the videotaping, and completes peer observation feedback form.

Phase VI. In this final phase the collaborative partners meet for the last time to share feedback and reflections from the revised lesson taught. They again engage in reflecting, rethinking, and assessing their lesson study work. Following this final meeting, each partner prepares a final written reflection to convey new understandings of students, teaching, learning, and their own professional development as a result of the lesson study process.

In this initial attempt of implementing the lesson study process in the field experience, preservice teachers were given the option of conducting the lesson study process either individually or with a collaborative partner. The preservice teachers were strongly encouraged to work in pairs and the benefits of shared work were explicitly discussed. In addition, there was a class session reserved for field experience work. However, there was concern that if the collaboration was forced then it might result in high levels of stress for the nontraditional preservice teachers as well as for those who travel long distances to attend the University. Those who selected to work independently completed the same process except they were required to teach both an initial lesson and a follow-up lesson. These students engage in a structured written reflection after each lesson and receive feedback from their cooperating teacher; however, they did not engage in peer collaboration.

Participants

Participants in this study were preservice teacher volunteers from one section of an elementary social studies methods class. There were a total of 25 class members with nine selecting to work independently on the alternative lesson study and sixteen selecting to work through the collaborative lesson study model. Eight of these preservice teachers agreed to participate in the study, four of which worked independently and four collaboratively on the lesson study process. All of the participants were white, female, traditionally college-aged elementary education majors. Six of the participants were first-semester seniors and two were second-semester juniors.

Data Sources and Analyses

Data sources used in the study were (a) individual interviews, (b) written reflections, and (c) instructor/ researcher field notes from in-class presentations and discussions. Data were collected in April and May of 2001. A graduate assistant, not associated with the class, conducted one face-to-face focused interview with each of the eight participants at the completion of the semester. Open-ended questions were asked such as; What did you find valuable about the lesson study process? In addition, each participant wrote five guided written reflection pieces that were collected at various points during the process. Field notes were recorded during the final two days of classes as students' presented and discussed their field experiences. The researcher field notes were used primarily for the purpose of triangulation. A series of multiple case studies (Yin, 1994) was selected as the most appropriate design for this study because of the nature of the research questions.

Tesch's (1990) systematic process of analyzing textual data was utilized to code the interview transcripts, field notes, and written journal entries into categories that allowed for the emergence of themes and patterns in the data. Construct validity (Yin, 1994) was addressed in this study through (a) using multiple sources of data (b) using different researchers to analyze the data, (c) establishing a chain of evidence, and (d) member checking. A summary report of the patterns of effects across the multiple cases studies is presented and organized around the two research questions in the next section.

Findings

How do Preservice Teachers Perceive the Lesson Study Model?

One of the strongest cross-case patterns identified was the value the preservice teachers found in talking with students about their thinking. All eight of the participants expressed an appreciation for gaining new understandings of the student perspective related to targeted social studies concepts during the planning phase. For example, one preservice teacher wrote:

> Prior to this field experience, I had never thought of the importance of questioning students to better understand their thinking related to concepts I was planning to teach. I was pleased with all the information they gave me and I had a much better understanding of their prior knowledge and misconceptions of life in the past.

In addition, several of the preservice teachers conveyed that they found the conversations with students so beneficial to their planning that they intend to continue with the practice. A participant stated in her interview, "I'm so glad that the student questions were included because I learned so much about the students' knowledge of the social studies content. I plan to ask children questions when I get ready to teach lessons in the future."

Six of the eight participants also felt that the gathering of professional literature related to the content or methodology of their lesson was beneficial to the planning phase.

A preservice teacher who worked on the project independently stated:

> The articles I found related to how elementary school students think about maps and their developmental abilities which was extremely useful in my planning. I thought previously that having to go look up articles related to the teaching or content we were planning was just a drag, not very useful, but now I see this differently. I found the articles greatly informed and changed the way I presented my lessons. I see value in reading professional literature because of this experience.

Although, six of the eight conveyed similar sentiments that researching the professional literature enhanced their knowledge and confidence in their selected content and/or methodology, one preservice teacher did not mention this as useful and one found it to be time consuming and not as rewarding as the other components in the process.

The value of observing the cooperating teacher as part of the lesson study research had mixed findings. Four of the eight stated that it was a somewhat to extremely beneficial opportunity for them as they prepared for their own instruction in the classroom. Gaining knowledge of the classroom procedures, types of typical student-teacher and student-student interactions, and focused attention to components of the lesson were cited as helpful to these four preservice teachers. The remaining four felt that they gained little from the experience and three were disappointed in the quality of the instruction observed. These three preservice teachers described lack of clarity, creativity, and connections to students' lives as areas of concern as they observed social studies lessons being taught.

The actual planning of instruction, implementation of the plan, and reflection on the lesson (Phases III-VI) was also considered valuable across each of the participants' cases. These participants were able to communicate both verbally and in writing their knowledge of the reflective practitioner role as an important part of their professional responsibilities. In addition to having a cognitive understanding of this professional responsibility, it appears through this experience they also begin to value the process in their own professional growth. For example, one participant shared in the interview, "I also learned how important the personal reflection is to teaching and it is something I want to continue to work on and to take seriously as I learn to teach and begin my practice." Another participant also emphasized how the process promoted professional growth through this written reflection:

> I think it is important as a professional to be able to self-assess and not to get discouraged when things do not turn out exactly as you had planned. I learned how to use my mistakes as a learning tool. When my partner taught the revised lesson it went much smoother. This experience gave me a chance to teach a lesson, reflect on it and make changes to it, and see how the revised lesson plan worked better.

Several of the participants discussed in their written reflections that this was the first time in their teacher preparation program that they had the opportunity to plan, implement, reflect, revise and then actually get to evaluate the effect the revisions made on the learning outcomes of the lesson. The processes of planning, observing, analyzing, reflecting, rethinking, and replanning involved in this experience were perceived as valuable across all of the cases examined in this study.

An additional pattern in the data that continued to emerge was that the preservice teachers gained a new appreciation for the time and energy required to teach effectively. One of the participants wrote, "I learned that teaching is a lot harder than I thought it was going to be, that it takes a lot of time and preparation, more than I previously thought it would." Similar statements occurred in five of the cases.

Although each of the participants provided evidence of unique understandings about teaching and content material as a result of their particular lesson study focus, there were no additional strong patterns that emerged across the cases. This is probably best explained by the fact that much of the preservice teachers' learning was lesson study specific. A good example of this is how one participant became aware of how important knowledge of students is when designing cooperative learning groups. From her research and experience this participant was able to write about the difficulty of implementing cooperative learning when you have little knowledge of the students. She experienced difficulty with classroom management during the initial lesson because she did not have adequate knowledge of the students' personalities and abilities in order to design effective groups. This new understanding was a result of selecting a cooperative learning strategy as a method of teaching her specific content. Others gained different information contingent on the nature of the content and their selection of methodology. This finding indicates that understandings gained about social studies content and pedagogy vary depending on the focus that is initially pursued and the direction that the participants follow as a result of reflection and analysis of the lessons taught.

What are the benefits and constraints of engaging in lesson study collaboratively as compared to working individually?

A clear distinction between those who worked collaboratively as compared to those who worked independently on this field experience was the repetition of terms used by the preservice teachers within the collaborative cases such as; teamwork, cooperation, peer feedback, interpersonal skills, and professional sharing and dialogue. Those who engaged in the lesson study model collaboratively consistently wrote and spoke of the value

and importance of working together with colleagues to improve teaching practices and learning outcomes for students. For example one collaborative partner wrote:

> I have learned through this experience that working cooperatively is essential in teaching. I know the value of constantly sharing ideas and resources and discussing the best ways to help students to learn the material. This experience helped me work on those interpersonal skills and cooperation skills that I plan to instill in my own students. We had to negotiate, listen to each other, remain flexible, and provide constructive peer feedback to improve the lesson we were designing. My partner brought her own strengths to the process and extended my thinking about both teaching and management practices.

There was no mention of new collaborative skills or appreciations for collaboration in the cases of preservice teachers who selected to engage in the independent lesson study. However, two of these preservice teachers documented that they would have preferred to engage in the process collaboratively if their complicated work schedules hadn't prohibited it. Another preservice teacher who worked independently offered these thoughts during the interview,

> I received very limited feedback from my cooperating teacher. And I think it would be a definite benefit to have a collaborative partner to work with, it would strengthen the opportunity to have dialogue and additional feedback and information about the lesson taught. I really missed out on this aspect.

When asked why they chose to work independently on the lesson study process they reported three main reasons. Three of the participants cited scheduling issues as the major constraint. Two of these preservice teachers were nontraditional students with families and one worked full-time; this made scheduling opportunities for the collaborative process very difficult. One of the participants was assigned to complete her field assignments in a surrounding county near her home. From the university this was a 45-minute drive; therefore, she was unable to identify a classmate that was willing to travel that distance to engage in the collaborative activities. Finally, one of the participants that had scheduling problems also stated that she preferred to work alone because she had bad experiences with collaborative projects in the past. She expressed that she felt it was easier to accomplish a task when she worked alone.

Discussion

Preservice teachers enter the classroom with a "critical lack of knowledge of pupils" (Kagan, 1992, p.1 42). Therefore, collaborative lesson study for prospective teachers should allow sufficient opportunity for inquiry into student thinking. The conversations with students were valuable for these participants because they gained a stronger knowledge of students' current conceptual understanding of the social studies content to be explored and appreciations for the importance of the student perspective to lesson planning. Assessing prior knowledge of students and connecting lessons with the student perspective are advocated characteristics of effective teaching (Brooks & Brooks, 1993).

Another benefit of this collaborative lesson study model is the explicit connection between attaining current understandings from the professional literature related to content and/or pedagogy and preparing instructional experiences. "If teachers investigate the effects of their teaching on students' learning, and if they read about what others have learned, they become sensitive to variation and more aware of what works for what purposes in what situations" (Darling-Hammond, 2003, p.279). Most of the participants of this study found the activity of researching the professional literature meaningful and relevant because they were able to use the knowledge gained to directly impact the instruction they would be implementing in the classroom. They felt a sense of purpose toward the research and experienced its usefulness to informed decision-making during the lesson planning process. However, there were exceptions to these sentiments among the participants. Some believe the amount of time and effort required to identify appropriate pieces of professional literature offsets the potential benefits to lesson planning. To assist preservice teachers in locating professional literature, it is suggested that they are introduced to and given reference lists of professional journals that will be particularly useful to informed decision-making during the lesson study planning process.

The final research requirement of observing social studies being taught in the classroom as a means of gathering additional data provides interesting findings. One half of the participants conveyed that the modeling provided was not useful to them as they planned their own social studies lesson. These participants stated that the social studies lessons observed consisted of students reading social studies material and responding to questions. Only one of these participants reported any follow-up discussion or extension activity to deepen student understanding. This highlights the inconsistency between the teaching methods advocated in the university and what the preservice teachers see practiced in the schools. This is particularly apparent in elementary social studies education because of the intense focus on the basics of literacy and math. There is often little time given to the planning and implementation of social studies lessons. In this case, the social studies methods professor must provide many opportunities for preservice teachers to experience effective social studies teaching through class and video demonstrations/modeling and

through controlled field observations.

These prospective teachers felt that the lesson study process (planning, teaching, reflecting, revising, reteaching, and reflecting again) was a rewarding professional development experience. They perceived this as a unique opportunity to actually be able to observe, analyze, and understand how the changes made to the initial lesson plan actually impacted the effectiveness of the lesson. According to Darling- Hammond (2003), "teachers need to be able to analyze and reflect on their practice, to assess the effects of their teaching, and to refine and improve their instruction" (p.278).

As preservice teachers move through teacher education programs it is not enough that they learn the terms reflective practitioner or teacher researcher; they need to develop related skills and positive dispositions that will be internalized into their vision of themselves as teachers. They must have models from which to operate and experience the usefulness of these models in their daily practice. Collaborative lesson study provides preservice teachers with a systematic strategy for improving teaching within their classrooms. These participants conveyed that they gained both skills and appreciations for these professional development processes; however, it is beyond the scope of this study to determine if they will actually become a part of their future practice.

Another benefit of the collaborative lesson study experience is that it clearly demonstrates for preservice teachers the complexity of teaching. These preservice teachers were amazed at the effort and time required to improve their teaching practices. This is an important realization because prospective teachers need to have a full awareness of the self-determination required of teacher professional development. However, if they experience the satisfaction of new understandings of teaching and learning as a result of the research, planning, analysis, and reflection, then hopefully these processes will be viewed as valuable professional roles and responsibilities that are worth the required time and effort.

The findings of this study also indicate that when preservice teachers engage in the lesson study model they gain knowledge and skills of collaboration and experience its value in enhancing professional understandings and practices. Langer (2000) identifies and describes six characteristics of teachers' professional lives that seem to make a difference in student learning. One of these characteristics is teacher participation in a variety of professional communities. Langer found that teachers who were effective at increasing student learning were not alone in their efforts they were connected with people with similar goals with whom they could plan and problem-solve. She writes:

> Teachers in the effective programs have at least one colleague at school, or someonewho taught elsewhere, or an interested significant other with whom to share joys, agonies, and ideas that affect instructional plans, decisions, and actions. They have contact with individuals who influence the way they view their subject, their students,and themselves as professionals. Through these interactions, they confront philosophical and superficial differences, learn from and challenge each other, and develop their own voices. (Langer, 2000, p.418)

In this study the preservice teachers who worked collaboratively talked extensively about how they gained new insights, perspectives, and collaborative skills as they worked with a peer through the collaborative lesson study process. There was evidence from each of the collaborative cases that the preservice teachers valued professional collaboration as a means to improve their pedagogical knowledge and practices. This was not found in the cases of those who worked independently on the lesson study process. These preservice teachers did receive feedback from their cooperating teachers and university professor but there was no indication that this was a collegial experience involving sustained professional planning, dialogue, and analysis. In fact, these preservice teachers communicated that they usually just do what the cooperating teacher suggests without question or dialogue because of their perception of their role within the mentor/mentee relationship. In contrast, those who worked through the lesson study process collaboratively received feedback from the university professor and cooperating teacher on their work; however, they also spent an average of four-six additional hours brainstorming, discussing, debating, and planning with a peer as they prepared and revised their lesson plan. This study demonstrates that the peer collaboration component of the lesson study model is valuable and if omitted then the process is diminished. It is recognized that additional support mechanisms need to be implemented so all preservice teachers can successfully engage in collaborative lesson study. Therefore, in future implementation the model will only be presented as a collaborative model. Additional class time will be dedicated to the process to insure adequate time and support is given to allow for successful collaboration.

Study Limitations and Future Research

Although I was gratified that participation in the modified lesson study model appeared to be a positive experience for the preservice teachers, I understand the limitations inherent in this study. For example, the selected research design of multiple case study leads to a small sample size and findings that are not generalizable. However, the description of how these perservice teachers experienced the lesson study experience provides valuable insight to teacher educators considering the use of this type of professional development model.

Another limitation of this study is that I was both the primary researcher and instructor of the course, which allows for bias. In addition, this study relied heavily on self-reports

from the participants who may have been trying to "please" me by expressing what they thought I wanted to hear. To address these limitations I recruited a graduate assistant not connected with the class to conduct the participant interviews after course grades had been assigned. The graduate assistant conveyed that all interview data would be kept confidential and verbally encouraged the participants to provide honest feedback to allow for improvement to the process. The graduate assistant also engaged in data analysis. We both coded the data and identified the patterns and themes that emerged independently then compared and discussed our findings. Both investigators identified the patterns and themes presented in this study.

Additional experimentation and research on lesson study as a professional development tool for preservice teacher education is needed. Studies that focus on understanding how lesson study affects preservice teachers pedagogical understandings and their conceptions of the roles and responsibilities of teachers, along with investigations into how to more fully support collaborative lesson study in different preservice teacher education models would be valuable.

Conclusion

Lesson study is an inquiry model of teacher professional development that supports active learning. The Holmes Group (1990) report asserts that perspective teachers should be engaged in thoughtful inquiry that encourages and promotes inquiry and reflection on practice. "From the time student teachers first begin seriously to hone their skills and to assume their professional attitudes, the habits of reflecting, questioning, and trying out and evaluating new ways of teaching - by themselves and with colleagues - should become embedded in their professional identity (Holmes Group, 1990, p.55).

Current professional development literature strongly advocates the need to foster a culture of support for teacher inquiry, examination, and sharing. Ball and Cohen (1999) assert that rich professional development experiences involve learning that: (a) centers around critical activities of teaching and learning such as planning lessons, evaluating student work, and developing curriculum; (b) grows from practical inquiry through cases, questions, analysis, and criticism; and (c) builds on substantial professional discourse that requires analysis and dialogue about practices and values. "These elements need to be part of a seamless process of professional learning that begins in preservice education, continues through the early years of induction, and extends through years of developing accomplished practice" (Darling-Hammond, 2003, p. 279).

The National Commission on Teaching and America's Future (1996) reports that reform efforts that invest time and resources in teacher learning and provide teachers with greater autonomy are the best hope for improving America's schools. Darling-Hammond and McLaughlin (1995) affirm that: "Professional development today means providing occasions for teachers to reflect critically on their practice and to fashion new knowledge and beliefs about content, pedagogy, and learners"(p.597). Collaborative lesson study embodies these ideals and is so extremely popular and highly valued by Japanese teachers that you will find it implemented in virtually every elementary and middle school in Japan (Stigler & Hiebert, 1999).

Why should American preservice teachers engage in the collaborative lesson study process? When a Japanese teacher was asked why she participated in lesson study she responded: "Why do we do research lessons? I don't think there are any laws [requiring it]. But if we didn't do research lessons, we wouldn't be teachers" (Lewis, 2000, p. 14). This statement indicates that the on-going, systematic analysis of teaching practices is imbedded in this Japanese teacher's conception of what it means to be a teacher. Ultimately, the continuous improvement of teaching is teacher's work; therefore, our teachers need not only active support from school administrators, but also a shared vision of how to go about this work and a set of professional values that will determine whether or not they will engage in the work. Providing early experiences in preservice teacher education with professional development models, such as collaborative lesson study, may allow our emerging teachers to develop professional identities that embrace new understandings of their roles and responsibilities as teachers. If we believe that real reform begins with *how* teachers teach, then we will need teachers who are equipped with the knowledge, skills, and dispositions necessary for the diligent work involved in the continuous improvement of teaching.

References

Ball, D. & D. Cohen. (1999). Developing practice,developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond and L.Sykes (Eds.), *The Heart of the Matter: Teaching as the Learning Profession*. San Francisco, CA: Jossey-Bass.

Brooks, J.G., & Brooks, M.G. (1993). *In search for understanding: The case for constructivist classrooms.* Alexandria, VA: ASCD.

Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership, 55*(5), 6-11.

Darling-Hammond, L. (2003). Teacher learning that supports student learning. In Ornstein, A., Behar-Horenstein, L.S., & Pajak, E. (Eds.), *Contemporary Issues in Curriculum* (pp. 277-282). Boston, MA: Pearson Education, Inc.

Darling-Hammond, L. & McLaughlin, M., (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 77* (7), 597-603.

Holmes Group (1990). *Tomorrow's schools: Principles for the design of professional development schools.* East Lansing, MI: The Holmes Group.

Kagan, D.M. (1992). Professional growth among preservice teachers. *Review of Educational Research, 62*(2), 129-

169.

Langer, J. A. (2000). Excellence in english in middle and high school: How teachers' professional lives support student achievement. *American Educational Research Journal 37,* 397-439.

Lewis, C. (2000, April). Lesson study: The core of Japanese professional development. Paper presented at the American Educational Research Association Annual Meeting, New Orleans.

Lewis, C. & Tsuchida, I. (1997). Planned educational change in Japan: The shift to student-centered elementary science. *Journal of Educational Policy 12,* 313-31.

Lewis, C. & Tsuchida, I. (1998). A lesson is like a swiftly flowing river: Research lessons and the improvement of Japanese education. *American Educator,* Winter, 14-17 & 50-52.

Lieberman, A., (1995). Practices that support development transforming conceptions of professional learning. *Phi Delta Kappan, 77* (7), 591-596.

National Center for Education Statistics. (1996). *Pursuing excellence: Initial findings from the Third International Mathematics and Science Study.* Washington, D.C.: U.S. Department of Education.

Stigler, J., & Hiebert, J., (1999). *The teaching gap.* New York, NY: The Free Press.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* New York: Falmer.Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological process.* Cambridge, MA: Harvard University Press.

Yin, R.K. (1994). *Case study research: Design and methods.* Thousand Oaks: Sage Publications.

Tracy C. Rock is an Assistant Professor of Elementary Education in the College of Education, University of North Carolina at Charlotte, NC 28223. Her research interests are: inquiry models of teacher professional development, elementary social studies education, and technology integration.

Frontiers in Research: Peace Education

David W. Johnson and Roger T. Johnson
University of Minnesota

The UNESCO constitution, written in 1945, states, "Since wars begin in the minds of men, it is in the minds of men that the defenses of peace must be constructed." This is a call for peace education, to educate all children and adults of the world in the values, philosophies, and competencies needed to build and maintain intrapersonal, interpersonal, intergroup, and international peace. The need for and importance of peace education seems to increase daily. The pervasiveness of psychological distress and pathology, anti-social behavior and dysfunctional relationships, prejudice and discrimination, and war and terrorism remind us daily of the costs of strife and dissention and the need for and difficulties in establishing and maintaining peace. Despite the need for peace and the costs of not living peacefully, however, very little agreement exists as to what should be taught in schools to help students learn how to create peace and how to keep it from deteriorating into destructive conflict, violence, and war.

A difficulty in trying to educate individuals in how to achieve inner peace and built peace among individuals, groups, and nations is deciding what is and is not peace education. Peace education has been defined as teaching about peace (its nature, how it may be established, reasons why it does not last), the philosophies underlying peace (such as nonviolence), the procedures and skills needed to implement the philosophies, violence and alternatives to violence, international issues such as the United Nations and nuclear weapons, environmental studies, and power and resource inequities. Peace education has also been defined as including social studies, history, and ethnic studies. Peace education has been defined as focusing on knowledge and understanding, competencies, attitudes and values, or personal characteristics such as optimism, self-regulation, and self-efficacy. The definition of peace education is so broad and inclusive that it is difficult to state what is *not* peace education, what are the parameters of peace education, or what are the characteristics of peace education.

The broad and somewhat amorphous definitions of peace education make it difficult for teachers to decide what to implement in their classrooms in order to provide peace education. The purpose of this section is to present four peace education programs that do not specifically fit within the definitions given above. Yet they all are vitally important for creating a peaceful world. They include acting in ethical ways, forgiveness, peer mediation, and decision making about difficult issues involving different positions and perspectives.

The first article focuses on ethical expertise and is authored by Darcia Narvaez, Rick Herbst, Scott Hagele, and Anna Gomberg. They present a program training students to act in ethical ways, a prerequisite for peace to be established and maintained. Acting ethically includes a sensitivity to what is and is not ethical, the reasoning about issues within the context of ethical principles, the motivation to act in ethical ways, and the ability to actually engage in ethical actions (i.e., actual engage in ethic behavior once it has been selected as the desired course of action). We should be striving to become more and more of an ethical person.

The second article, authored by Robert Enright, Elizabeth Gassin and Jeannette Knutson, focuses on the need for forgiveness in order to live peacefully. Violence and other forms of destructive conflict often result in anger and resentment which in turn may result in retaliation and other actions that escalate the conflict in destructive and negative ways. An essential competence for creating peace is forgiveness. Even in the most difficult situations, such as in Belfast schools in Northern Ireland, students may be taught forgiveness as a way to counter their anger, anxiety, and depression resulting from their exposure to violence. Enright and his colleagues note that children in Northern Ireland represent a unique risk group, as they are exposed to poverty and both acute and chronic forms of violence. They are implementing a forgiveness curriculum in the Northern Ireland schools to counteract the effects of violence and terrorism in the children's lives.

One of the unresolved issues in peace education is the age at which it can effectively begin. Stevahn and Oberle demonstrate that effective conflict resolution training can be successfully implemented in kindergarten, an emphasis on perspective reversal (i.e., taking the perspective of the other disputant) increased the effectiveness of conflict resolution, kindergarten children were able to coordinate their perspective with the perspective of the other disputant, and embedding the conflict resolution training in an academic unit increased the children's academic achievement. Effective peace education, therefore, may begin at a very young age.

One issue with peace education is whether it can be taught widely throughout the schools. An example of a widescale implementation of constructive controversy as part of a leadership and citizenship curriculum is provided by James Mitchell and his associates. This program was taught to 1,368 K-12 students in an inner-city school district over an eight-week period. The success of this project may increase the optimism of peace educators about the potential of what they do.

The final article, authored by David and Roger Johnson, notes that in order to establish and maintain peace, it is necessary to discuss difficult issues (such as ethnic or religious differences, economic interests, and whether all members of

a society are of equal worth) and come to some accommodation to the differences of positions and perspectives that exist among the relevant parties. They present a meta-analysis of twenty-eight studies on implementing the academic controversy procedure within schools, involving students from primary, intermediate, secondary, college, and adult training levels. This program trains students in how to research and develop a position, present a persuasive case for their position, refute other positions while rebutting attacks on their own, reversing perspectives to see the issue from all sides, and (dropping all advocacy) come to their best reasoned judgment about what should be done. When enough students have been taught how to engage in the controversy procedure, a tipping point may be reached so that difficult issues may be discussed in establishing and maintaining peace.

To build and maintain peace, children, adolescents, and young adults need to learn how to act ethically, how to forgive, how to resolve conflicts constructively, and how to engage in intellectual arguments over difficult issues. Each of these programs is vitally important for educating individuals how to create and maintain peace.

Nurturing Peaceful Character

Darcia Narvaez, Rick Herbst, Scott Hagele, & Anna Gomberg
University of Notre Dame

In this article, we briefly describe a model for character education that offers a foundation for structuring peace education. The Ethical Expertise model (Anderson, Narvaez, Bock, Endicott, & Lies, 2003; Narvaez, 2003; Narvaez, Endicott, & Bock, in press) is the framework underlying character development education (CDE), the integration of the ancient Greek understanding of moral character with cognitive developmental psychology and pedagogy (Narvaez, in press; 2003). Similar to Plato's notion of the moral person as one with highly developed skills (techne), the EthEx model provides a set of skills that comprise ethical know-how and lead to human flourishing (Aristotle's eudamonia). The ExEth model of skills, developed under the auspices of the Minnesota Community Voices and Character Education project identifies four components or psychological processes that are required for ethical action to ensue: ethical sensitivity, ethical judgment, ethical motivation, and ethical action. Each process includes a set of teachable skills, including skills for getting along with others, making moral decisions, living the good life and self-actualization.

If there were a widespread comprehensive approach to peace education in the country, would the United States have gone to war with Iraq in 2003? If the populace had the habit of finding peaceable solutions to conflicts and respecting others who have different worldviews, would the option of war have seemed as 'logical'? We think not. If citizens were oriented to peaceable living, there would have been much more deliberation about the means for ousting Saddam Hussein and about the consequences of war. The fact that our leaders displayed no interest in hearing viewpoints other than hawkish ones and the populace did not demand it underscores the need for peace education.

Peace education ought to be a primary component of every approach to moral education and character development. After all, living peacefully is a fundamental goal of the moral life. In a search for programs and models of peace education we found few and so it appears that there is a gap between what needs to be done and what is being done.

In this article, we briefly describe a model for character education that offers a foundation for structuring peace education. The questions we address include: What are the skills for peaceful living? How does one develop skills for peaceful living? How should peaceful know-how be taught? We also briefly compare two classroom approaches with our approach, discuss the violent-media-saturated environment that surrounds children these today, and identify several web-based resources.

The Ethical Expertise Model for Character Development Education (EthEx)

The Ethical Expertise model (EthEx) (described in Anderson, Narvaez, Bock, Endicott, & Lies, 2003; Narvaez, 2003; Narvaez, Endicott, & Bock, in press) is the framework underlying character development education (CDE), the integration of the ancient Greek understanding of moral character with cognitive developmental psychology and pedagogy (Narvaez, in press; 2003). Similar to Plato's notion of the moral person as one with highly developed skills (techne), the EthEx model provides a set of skills that comprise ethical know-how and lead to human flourishing (Aristotle's eudamonia). The skills for ethical know-how are based on a broad review of literature and fall into the categories established by Rest (1983) in his four-component model of moral behavior. The four components are the psychological processes that must have occurred for an ethical action to ensue and they are: ethical sensitivity, ethical judgment, ethical motivation, and ethical action. The ExEth model of skills, developed under the auspices of the Minnesota Community Voices and Character Education project (Anderson et al, 2003), divides the four components into teachable skills, including skills for getting along with others, making moral decisions, living the good life and self-actualization.

What are the skills for peaceful living?

The ancient Greeks believed that everything about a

life impinges on moral character, that moral growth is a lifelong endeavor, and that persons must constantly self-monitor their choices and actions. The EthEx skills are character skills that should be developed by everyone across situations and cultures. (For more on the model and skills see sources cited above.)

Here we highlight the EthEx skills that are particularly related to peaceable living. For the full list of EthEx skills (and suggestions for subskills), see Table 1. Ethical know-how includes skills for living in peace with others that cut across ethical sensitivity, ethical judgment, ethical motivation and ethical action.

Ethical sensitivity is the ability to notice, interpret and respond to stimuli and the behavior of others according to moral conceptual structures (schemas). For example, an experienced teacher is able to notice when student behavior is just beginning to get out of hand whereas a novice teacher does not notice until behavior is well out of hand (Clarridge & Berliner, 1991). An experienced teacher also has an understanding of how her routines or a lack of routines affect student behavior. Some of the skills more important to moral sensitivity include perspective taking (Selman, 2003), managing anger and aggression (Dodge & Tomlin, 1987; Gibbs, Potter, Barriga, & Liau, 1996) and controlling social

Table 1

EthEx Skills: Four Processes, Their Skills and Suggested Subskills

Ethical Sensitivity	Ethical Judgment
ES-1: Understand Emotional Expression	EJ-1: Reasoning generally
Identify and express emotions	Reasoning objectively
Finetune your emotions	Use sound reasoning & avoid reasoning pitfalls
Manage anger and aggression	Make scientific method intuitive
ES-2: Take the Perspectives of Others	EJ-2: Reasoning ethically
Take an alternative perspective	Judging perspectives
Take a cultural perspective	Reason about standards and ideals
Take a justice perspective	Reason about actions & outcomes
ES-3: Connecting to Others	EJ-3: Understanding Ethical Problems
Relate to others	Gathering information
Show Care	Categorizing problems
Be a Friend	Analyzing ethical problems
ES-4: Responding to diversity	EJ-4: Using Codes and Identifying Judgment Criteria
Work with group and individual differences	Characterizing codes
Perceive diversity	Discerning code application
Become multicultural	Judging code validity
ES-5: Controlling Social Bias	EJ-5: Understand consequences
Diagnose bias	Choose your environments
Catch stereotyping & overcome automatic responses	Predicting consequences
Nurture Tolerance	Responding to consequences
ES-6: Interpreting situations	EJ-6: Reflect on the Process and Outcome
Determine what is happening	Reasoning about means and ends
Perceive morality	Making right choices
Respond creatively	Monitoring one's reasoning
ES-7: Communicate Well	EJ-7: Coping
Speak and listen	Apply positive reasoning
Communicate non-verbally and alternatively	Managing disappointment & failur
Monitor communication	Developing resilience

Table 1 (continued)

EthEx Skills: Four Processes, Their Skills and Suggested Subskills

Ethical Motivation	Ethical Action
EM-1: Respecting Others	EA-1: Resolving Conflicts and Problems
Be civil and courteous	Solve interpersonal problems
Be non-violent	Negotiate
Show reverence	Make amends
EM-2:Cultivate Conscience	EA-2: Assert Respectfully
Self command	Attend to human needs
Manage influence & power	Build assertiveness skills
Be honorable	Use rhetoric respectfully
EM-3: Act Responsibly	EA-3: Taking Initiative as a Leader
Meet obligations	Be a leader
Be a good steward	Take initiative for and with others
Be a global citizen	Mentor others
EM-4: Help Others	EA-4: Planning to Implement Decisions
Cooperate	Thinking strategically
Act thoughtfully	Implement successfully
Share resources	Determine resource use
EM-5: Finding Meaning in Life	EA-5: Cultivate Courage
Center yourself	Manage fear
Cultivate commitment	Stand up under pressure
Cultivate wonder	Managing change and uncertainty
EM-6: Valuing Traditions and Institutions	EA-6: Persevering
Identify and value traditions	Be steadfast
Understand social structures	Overcome obstacles
Practice democracy	Build competence
EM-7: Develop Ethical Identity & Integrity	EA-7: Work Hard
Choose good values	Set reachable goals
Build your Identity	Manage time
Reach for Your Potential	Take charge of your life

bias (Singh, 1991). For example, perspective-taking is important for interpersonal and intercultural relationships and is positively correlated with prosocial behavior (Eisenberg & Mussen, 1989). In order to be moral people, students need to learn to take the perspectives of other cultural groups. Students also need to learn to take a justice perspective by, for example, filtering societal events from perspectives of the least fortunate and most voiceless. Perspective-taking skills underlie higher level skills in moral reasoning (Kohlberg, 1984).

Ethical judgment involves reasoning about means and ends in light of principles (ethical codes) and context. Those with more complex, postconventional moral reasoning are less likely to harm others (Kohlberg, 1984) and more likely to act according to their principles (Rest, 1986; Rest & Narvaez, 1994). In order to develop moral reasoning, students need to discuss moral conflicts and dilemmas, particularly with peers who have different perspectives (Kohlberg, 1984). Persons who seek out intellectually stimulating environments are more likely to develop higher levels of moral reasoning (Rest,

1986). Multicultural experience can positively affect moral reasoning development as well as intercultural skill development (Endicott, Bock, & Narvaez, 2003). Ethical judgment includes not only moral reasoning development but also the skills of cognitive control—learning to control, balance and guide one's reasoning through such things as the management of disappointment and failure, coping strategies and resiliency (Wang, Haertel, & Walberg, 1998). For example, optimistic thinking is a skill that must be continually practiced in what is considered by many to be a pessimistic world. Type of thinking (optimistic or pessimistic) affects one's belief structures and judgments, attributions toward others, perspectives on the future, and perseverance under adverse conditions (Seligman, 1995).

Ethical Motivation involves prioritizing ethical action over other personal needs and desires. Moral exemplars cultivate a moral identity that orients them to help others flourish (Colby & Damon, 1991). Ethical motivation skills that moral exemplars have include being respectful, showing reverence, centering oneself, cultivating wonder, and nurturing wisdom. All are central to moral identity (Dalai Lama, 1999). Respect for persons stems from the belief that all persons, including the self, have value, rights and responsibilities. Respect includes following rules of courtesy and civility, and respecting the dignity of life in general. For example, using language in the classroom that confirms others increases the likelihood that student decision making includes the perspectives and inputs of others (Johnson and Johnson, 1989). Fundamental to being respectful is a sense of reverence, not only for creation but for self. The ancient Greeks understood humans to have potential excellences and the aspiration to actualize them, requiring the deference of others and supportive political-social conditions for their unfolding. Religious notions of human beings as created in the image of God, each as a temple for the Holy Spirit, provide a sense of what reverence for self and others means. Reverence for others does not mean narcissistic self-aggrandizement, which is sometimes mistakenly promoted in school-based self-esteem programs.

Adults need to coach children on building a self concept as an ethical person. Adults help children build an ethical identity by referring to a child's good nature when they are being helpful (i.e., "You are a good boy") and as the source of a moral action (e.g., "You are a good boy. Helping your sister pick up the toys is what good boys do") (Grusec, Goodnow, & Cohen, 1991). Children need to learn to adhere to ethical codes, both professional and personal (e.g., national codes of democratic values and tolerance of difference, family-community religious codes). In order for children to be able to learn the skills of cooperation, acting thoughtfully, and sharing resources, they have to feel cared for and safe (Watson & Eckert, 2003). This allows them to take interest in the welfare of others. An interest in maintaining social harmony for the welfare of all motivates peace-making, cooperation, and hopefulness (Dalai Lama, 1999). Participation in helping others promotes a helping attitude (Staub, 1978).

Ethical Action is comprised of the skills necessary to complete a moral action once it has been selected. Although the other skills (sensitivity, judgment, motivation) may be chronologically prior to implementing ethical behavior, they are not sufficient to produce the behavior. In other words, it is not enough to notice (sensitivity), to judge what action is most moral, and to be motivated to take it. One must also know how to carry out the moral action. Moral action often fails because the individual does not know what steps to take or loses confidence. Students need to practice many ways of performing many kinds of ethical action. Students need to learn supporting skills such as attending to human needs and managing fear. In order to maintain courage to execute and act on their beliefs, children need to learn to stand up under pressure, and manage change and uncertainty. Further, conflict resolution and negotiation skills are necessary for successful social interaction (Lantieri & Patti, 1996). Children who get along well with others, resolving conflicts easily, are more helpful and altruistic (Eisenberg & Mussen, 1989).

An additional skill, assertiveness, usually left out of traditional character education programs, is related to positive interpersonal behavior (Eisenberg & Mussen, 1989). Assertive behavior is related to self-regulation. It involves choosing for oneself (not for others and not others for self); it is expressive (not inhibited, and not depreciative of others); self-enhancing (but not at the expense of another); and can achieve the desired goal (but not by hurting others). Assertiveness skills enable one to solve problems, resolve conflicts, and help prevent depression (Seligman, 1995). Children need to be coached on assertiveness. They need to learn how to encourage themselves when tasks get difficult and when obstacles arise to completing moral goals. They need to learn how to distract themselves from temptations and to seek support from others when completing an ethical action.

The setting: How does one develop skills for peaceful living?

Based on review of literature on how expertise is developed, the EthEx approach identifies three key features of settings that nurture ethical character: well-structured environments, coaching of skills, and student self-regulation. First, adults need to set up well-structured environments that teach appropriate ethical intuitions (Hogarth, 2001). Too often, adults do not think about what lessons are being taught by, for example, the recurrent patterns of reward and punishment in classrooms. Too often, schools have stifling and unsafe environments that discourage moral behavior. The climate and reward structures feature strongly in promoting ethical character development. For example, teachers should expect students to treat each other with respect, explicitly discuss the benefits of prosocial action for others, promote an atmosphere of fairness and equality, establish the classroom as a considerate and sensitive community, and model and encourage forgiveness as a

helpful and healthy response. Students need to feel a sense of belonging and a sense of community in order to flourish and be motivated (Solomon, Watson & Battistich, 2002).

Second, adults need to design instruction that moves students from ignorance to competence in ethical know-how using a notice-to-expert approach (Marshall, 1999). What do moral experts know? How are they different from novices? Moral experts have more developed skills in the moral excellences, social, personal, and citizenship skills. Moral experts in a particular excellence have more and better organized knowledge about it, have highly tuned perceptual skills for it, and have highly automatized and effortless responses. Too frequently, character education programs rely on methods that are anathema in academic instruction (because they don't work) such as exhortation, rewards, and rote memorization. It is not enough to present information, tell students about things and have them complete a worksheet (i.e., the "receptive-accrual" approach in which the teacher 'pitches' information and the student 'catches' it or is stupid, Anderson, 1998). There needs to be intentional stages of skilled coaching. Adults need to provide extensive, coached practice in ethical skills. Students, like any developing expert, need multiple opportunities to practice in a variety of contexts with close guidance.

The Minnesota "Community Voices and Character Education" project suggests four levels of novice-to-expert instruction. Teachers should begin with the level their students require and move to the next level when students are prepared. In Table 2 we have provided an example for developing skills of peacemaking using the four levels of expertise development. In Level 1, students need to be immersed in examples and opportunities. They must experience peacefulness, its effects, its meaning, and its practice. During immersion experiences, the teacher helps the student learn to see the 'big picture' and recognize basic patterns. In Level 2, students learn to pay attention to important features by studying exemplars and how they perform. Students are given many experiences that help them build more and more elaborative knowledge about the domain. In Level 3, students practice setting goals, plan the steps to solve problems, and use sets of skills in concert. In Level 4, students integrate their knowledge and procedural skills as they practice in many contexts and execute problem solving strategies.

Third, students need to be coached to develop self-regulation skills so that they can monitor their own character development. Self-regulation increases achievement in academics (Zimmerman, Bonnner, & Kovach, 2002). We expect the same for character development. Within an immersion experience that trains the appropriate intuitions, students should be given the autonomy to make choices. Educators should help children use their deliberative, conscious minds to guide select environments and make choices that nurture appropriate moral intuitions. Students must be empowered to monitor their own character development by using skills to address and continually answer the question: Who should I be?

Peace education in a violent, materialist culture

The three features of setting critical to peace and ethical education are well-structured environments that teach appropriate ethical intuitions, coaching of ethical skills, and student self-regulation of character. The ubiquitous nature of media allows its components (e.g. motion pictures, television, Internet) to have a major role in the formation of children's environments. Due to the powerful interplay between children's environments and media, both positive and negative elements of media should be discussed when assessing peace and ethical education. While the positive aspects of media will be discussed in relation to the Internet, the daunting negative features of media will be addressed first. If one examines children's lives in the U.S. today, one can fault them on all counts. The environments in which most children in the U.S. find themselves are not well-structured for ethical development. On the contrary, they are well-structured for learning selfishness and violence. Advertisers pummel children with their products in every venue they can reach, including the classroom. Television increasingly exposes children to violence and self-gratification at all hours of the day on virtually every channel. In a way the media is coaching children on being selfish and violent and teaching them to self-regulate on becoming physically and materially attractive to others.

Violence in the media is an increasing concern. According to the American Psychological Association website (August, 2003), "viewing violence on the screen" affects children in the following ways: It increases self-protection and mistrust of others because of an increase in fear of being victimized. Viewing violence on the screen decreases sensitivity to violence and the likelihood of intervening on behalf of a victim. It increases an appetite for violence. Sexual violence depictions increase male aggression against females. Derksen and Strasberger (1996) say that "the major effects of violence, antisocial behavior, and aggression in the media can be placed into the following categories: 1) displacement of healthy activities, 2) modeling of inappropriate behavior, 3) disinhibition, 4) desensitization, 5) aggressive arousal and 6) association with risk taking behavior." Huessman et al (2003) note the power of watching violent heroes on interpersonal behavior decades later. Their data show that watching violent TV shows is like smoking cigarettes. The more cigarettes you smoke (the more violent shows you watch), the more likely you are to develop lung cancer (the more likely you are to be aggressive in your interpersonal relationships).

Anti-violence education programs include the EQUIP program (Gibbs, Potter & Goldstein, 1995) which involves building social and coping skills, as many anti-violence programs do, as well as moral reasoning skills. Such programs can be successful at preventing further criminal physical violent behavior. Yet this is far from peace education. One must distinguish anti-violence education, which is often focused primarily on the individual and on prevention, from peace education, which requires a proactive orientation

Table 2

Ideas for Developing Peacemaking Skills using the EthEx Framework

Real-Life Exemplar: Former President Jimmy Carter is one of the great peacemakers of our time. As President of the United States he concerned himself with human rights and mediating peace where necessary, most notably, between Israel and Egypt. The Carter Center, which he founded and continues to animate, mediates present-day conflicts throughout the world.

Level 1: Immersion in Examples and Opportunities

Attend to the big picture, Learn to recognize basic patterns

What is peacemaking? Look at sources from the library and on the web about peacemaking to discover how people define it. Discuss with the class possible definitions that you revisit over the course of the unit, finalizing a definition at the end.

What do peacemakers do? Show examples of socio-political peacemaking in the world (e.g., tearing down of the Berlin wall, Nelson Mandela and the reorganization of South Africa). Find examples of peacemakers in the local community (e.g., court mediators) to bring. Discuss who have been peacemakers in their lives, families, and so on. Students reflect on the characteristics of a peacemaker.

Peacemaking in action. Invite peace activists or others who promote peace in everyday life (e.g., Buddhists) to speak to the class about their traditions, what they do and why. Have students identify similarities and differences with their own traditions.

How do we maintain peace in our lives? Have the students interview family members (or elders in the community) about how peace is maintained in their homes, e.g., rules, chores, honoring parents, etc. Who intervenes in conflicts at home? At school? Who among their friends would they describe as a peacemaker? Why?

Level 2: Attention to Facts and Skills

Focus on detail and prototypical examples, Build knowledge

Peacemaking in everyday life. For a week, students identify peacemaking opportunities in television stories, books, their own experiences. Assess with a report on what they've observed.

Conduct research on programs that promote peacemaking methods. Investigate negotiating and conflict resolution skills that might be used to facilitate peacemaking. Have the students seek out different programs which might further peacemaking. How might they apply to problems in relationships with others? With family?

Investigate steps toward peacemaking. Encourage the students to brainstorm about the steps necessary to bring peace to a contentious situation, e.g. bring the parties together, insure that they listen to one another, avoid expressions of anger and resentment, etc.

Peace in literature and poetry. Have the students seek out works of literature or poetry which highlight the importance and the virtue of peace. Give them the opportunity to share it (or a portion of it) with the class and explain why they chose that particular work.

Investigate the role of law in maintaining peace. Students should consider how laws, from local to international, further the maintenance of peace in our local communities, our nation and the world.

Table 2 (continued)

Ideas for Developing Peacemaking Skills Using the EthEx Framework

Invite guests who have experienced conflict. Have the students invite exchange students (from countries that have experienced conflict or struggle) or veterans to speak of their experience of war or conflict. Ask the guests to reflect on the value of peace and the importance of peacemaking. Have the students write an essay on the importance of being peacemakers.

What are the habits of people who live peaceably? Have students study the details of the lives of peace makers (e.g., the Dalai Lama) and make a list of their practices (e.g., they manage their emotions and appetites, look at the positive side of things, maintain an empathic perspective towards everyone, treat others as equals, forgive those who hurt them, make amends with those they have offended, stay "centered").

Level 3: Practice Procedures

Set goals, Plan steps of problem solving, Practice skills

Examine those places where peacemaking is needed. Have the students find examples of places, situations, domestic and international conflicts, where peacemaking is needed. Analyze what is the root of the problem in each case, creating a cognitive map for each perspective. Brainstorm on what can be done to remedy the contentious, perhaps violent, situation?

What brings about the opposite of peace? Examine world history to find evidence of conditions that existed prior to a war breaking out. What human needs were not being met? What could have been done to avert the war? What can we do now to avert war in areas of conflict?

Class meetings. Use class meetings as a vehicle for practicing peaceful resolution strategies. These can be regularly scheduled and/or held when a conflict arises. Unless students are skilled in self-government, the class meetings should be led by teachers. For best results, they should last usually two minutes long on a consistent basis.

Practice using peacemaking statements. Students practice saying peacemaking statements that they can use when conflict arises around them. Students identify opportunities to use them. Statements include: "Take it easy," "It's no big deal," "Let's go," "Cool it," etc.

Practice the habits of peaceful living. After discussing the related activity in Level 2: "What are the habits of people who live peaceably?," ask students to select one of the habits to practice developing over a week or month's time. Have them keep a journal to report on their progress (recording when they made the effort and how it went).

Level 4: Integrate Knowledge and Procedures

Execute plans, Solve problems

Using peacemaking at school. After identifying school situations in which peacemaking is difficult to maintain, students identify one type of situation in which they will practice (individually or as a group) peacemaking during the week. Practice and report on the outcome, difficulties and successes.

Fostering peace through pen pals. Give the students the opportunity to form pen pal relationships with students from other countries and cultures. Encourage discussion about how communicating with others can foster ongoing peace between individuals and nations.

Table 2 (continued)

Ideas for Developing Peacemaking Skills Using the EthEx Framework

Sample Student Self-regulation Items for Peacemaking

I try to end fights, not start them.
I talk through my conflicts with others in a respectful manner.
I try to find ways to help other people solve conflicts.
I notice what behaviors hurt other people and avoid those behaviors.
I know how to work with others in a group:
- I take turns talking.
- I take turns being leader.
- I take turns being the recorder.
- I listen to others.

towards cooperation and a positive focus on others and community.

We found two approaches to peace education that we briefly compare to EthEx: Timpson's *Teaching and learning peace* and Lantieri and Patti's *Waging Peace in our schools*. Both have features that make them attractive for classroom usage. Timpson (2002) contends that peacemaking can be taught if there is the will to do so. His book primarily addresses ethical sensitivity and ethical judgment. The skills pertinent to developing ethical sensitivity fall neatly into Timpson's model for teaching and learning peace. Here are some examples of overlap. Ethical Sensitivity (ES) 1: Timpson argues that for peacemaking to occur, *emotional intelligence* is equally important, if not more important than normal intelligence. ES 2: *Taking the perspectives* of others promotes peace because it allows one to better understand why an individual acts in a certain way. By understanding another's thought processes, one is more likely to generate peaceful solutions to conflicts. ES 3: Timpson discusses the importance of *connecting to others* in working on problem solving and group consensus. Establishing a sense of connection is an important step towards reaching an effective and peaceful end. ES 4: Timpson argues that *diversity* should be used to prompt educators to show their students how all people are connected, even, for example, how one person's privilege might be predicated on another person's suffering. ES 5: Timpson encourages students and educators to *control social bias* by reassessing one's and being committed to those things that are life affirming. ES 6: *Interpreting situations* is a vital component of emotional intelligence. ES 7: *Communicating well* is an important tool for the peacemaking process. Communication is the essence of a three-principle model for achieving peace—listening, expressing empathy, and seeking consensus. Like EthEx, Timpson encourages studying and imitating role models. However, the Ethex model provides a broader view than Timpson's in that it nurtures not only ethical sensitivity and reasoning skills, but also ethical identity and the skills necessary to implement peacemaking. Ethical behavior of any kind, including peacemaking, is a set of processes for which multiple skills are needed.

In *Waging peace in our schools*, Lantieri and Patti (1996) discuss the peacemaking curricula of the Resolving Conflicts Creatively National Center designed to develop social and emotional competencies in students of all ages. The center promotes peaceable schools and classrooms through a comprehensive program that includes a selected set of skills that cut across the EthEx skills: conflict resolution (EA 1), cooperation (EM 4), caring (ES 3), appreciation of diversity (ES 4), responsible decision making (all of ethical judgment), and appropriate expression of feelings (ES 1). They emphasize the importance of adults coaching these skills as students practice them across a variety of contexts. Students learn to give "I" messages, listen actively, peer mediate, and become culturally competent. This approach fits nicely with EthEx in using a similar pedagogy and advocating several of the skills identified in the EthEx model.

Peace Education Resources

While media often is justifiably criticized for its violent nature, one should acknowledge media sources that critique and confront violence in the media and the world. While exemplars exist in other media, we have identified three websites that move against trends to depict aggression positively. In particular, these websites are useful for teachers promoting peaceful students and offer information and entertainment directly to students.

(1) http://www.cultivatingpeace.ca/pematerials/: This website provides lesson plans, discussion questions and other peace resources for educators. It also houses suggestions from other educators about teacher behavior,

for example, how teachers should behave towards students to promote peacefulness and how to lead student discussion and decision making. Through the website students have the opportunity to network with other students to discuss peace and related issues. An important component of this website is the "peacemaking challenge" quiz that allows students to combine information gained from classroom activities and employ that information. The quiz draws attention to the proper and improper actions to take in situations that challenge peace.

(2)http://www.haguepeace.org/index.php?name=education: This site offers tools for educators of all ages, from primary to adult. Not only are there lesson plans that match well with EthEx skills, the site provides a theoretical literature review. One classroom activity that fosters the abilities necessary in the presentation and acceptance of contrasting opinions is the following: Students write an opinion (that they think only they hold in the group) on a placard and tape it to themselves. Students then interact with classmates in order to find other individual of a similar persuasion. Groups are then asked to explain why an opinion is held and actively listen to rationalizations by other groups. This activity, similar to other activities on the website, builds skills across component areas—sensitivity, judgment, action.

(3) http://www.peace.ca/peaceeducationtheory.htm: This website offers a theoretical perspective on peace education. The orientation presented is compatible with character development education and EthEx skill development. For example, the site says: "Peace education is based on a philosophy that teaches nonviolence, love, compassion, trust, fairness, cooperation and reverence for the human family and all life on our planet. Skills include communication, listening, understanding different perspectives, cooperation, problem solving, critical thinking, decision making, conflict resolution, and social responsibility.

Conclusion

A culture that highlights and glorifies violence in its media, in what it presents to children and youth, and in how it approaches world conflict, peace education is deeply needed. But a movement for peace education is challenged by a citizenry that has grown desensitized from being immersed in such a culture. It is vital that parents and children develop media literacy in order to combat desensitization and to discourage even subtle violent attitudes and behaviors in themselves and others.

Peace education needs a guiding theory, and we believe that character development education provides it by focusing on human flourishing. EthEx skills offer concrete ideas about what should be taught. The four components of ethical sensitivity, ethical judgment, ethical motivation and ethical action provide a process framework for thinking about peaceful action. EthEx offers specific suggestions for establishing environments that are structured for ethical development. EthEx encourages the use of novice-to-expert pedagogy and the extensive coaching of skills. EthEx stresses the importance of student self-regulation in the successful cultivation and maintenance of ethical character.

References

Anderson, C., Narvaez, D., Bock, T., Endicott, L., & Lies, J. (2003). *Minnesota Community Voices and Character Education: Final evaluation report.* Roseville, MN: Minnesota Department of Education.

Anderson, L.M. (1989). *Learners and learning.* In M.C. Reynolds (Ed.), Knowledge base for the beginning teacher (pp. 85-99). Oxford: Pergamon Press\Anderson, L.M. (1989b).

Clarridge, P. B., & Berliner, D. C. (1991). Perceptions of student behavior as a function of expertise. *Journal of Classroom Interaction, 26,* 1-8.

Colby, A., & Damon, W. (1992). *Some do care.* New York: Free Press.

Dalai Lama, (1999). *Ethics for the new millenium.* New York: Riverhead Books.

Derksen, D.J & Strasburger, V. (1996). Media and television violence: Effects on violence, aggression, and antisocial behaviors in children. In A.M. Hoffman (Ed.), *Schools, violence, and society* (pp. 61-77). Westport, CT: Praeger Publishers/Greenwood Publishing Group.

Dodge, K.A., & Tomlin, A.M. (1987). Utilization of self-schemas as a mechanism of interpretational bias in aggressive children. *Social Cognition, 5,* 280-300.

Eisenberg, N. & Mussen, P. H. (1995). *The roots of prosocial behavior in children.* Cambridge, MA: Cambridge University Press.

Endicott, L., Bock, T., & Narvaez, D. (2003). Moral Reasoning, Intercultural Development, and Multicultural Experiences: Relations and Cognitive Underpinnings. *International Journal of Intercultural Relations, 27,* 403-419.

Gibbs, J.C., Potter, G.B, & Goldstein, A. P. (1995). *The EQUIP program: Teaching youth to think and act responsibly through a peer-helping approach.* Champaign, IL: Research Press.

Gibbs, J.C., Potter, G.B., Barriga, A.Q., & Liau, A.K. (1996). Developing the helping skills and prosocial motivation of aggressive adolescents in peer-group programs. *Aggression and Violent Behavior, 1,* 283-305.

Grusec, J. E., Goodnow, J. J., & Cohen, L. (1996). Household work and the development of concern for others. *Developmental Psychology, 32,* 999-1007.

Hogarth, R. M. (2001). *Educating Intuition.* Chicago: University of Chicago Press.

Johnson, D. W., and Johnson, R. T. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Kohlberg, L. (1984). *The Psychology of moral development: Essays on moral development, Vol. II.* New York: Harper

& Row.
Lantieri, L., & Patti, J. (1996). *Waging peace in our schools.* New York: Beacon Press.
Marshall, S.P. (1995). *Schemas in Problem Solving.* Cambridge, Cambridge University Press.
Narvaez, D. (in press). Character development education: Justice and virtue, character and community. In M. Killen & J. Smetana (Eds.), *Handbook of moral development.*
Narvaez, D. (2003). *Educating moral intuition.* Manuscript in preparation.
Narvaez, D., Bock, T., & Endicott, L. (in press). Who should I become? Citizenship, Goodness, Human Flourishing, and Ethical Expertise. In W. Veugelers and F. Oser, *The positive and negative in moral education.* New York: Peter Lang.
Palmer, P. (1999). The grace of great things: Reclaiming the sacred in knowing, teaching, and learning. In S. Glazer (Ed.), *The heart of learning: Spirituality in education* (pp. 15-32). New York: Tarcher.
Rest, J.R. (1983). Morality. In P. Mussen (Gen. Ed.) *Manual of Child Psychology*, J. Flavell & E. Markham (Vol. Eds.) Vol.3: *Cognitive development* (pp. 556-629). New York: Wiley.
Rest, J.R. (1986). *Moral judgment development: Advances in research & theory.* New York: Praeger.
Rest, J.R. & Narvaez, D. (Eds.) (1994). *Moral development in the professions: Psychology and applied ethics.* Hillsdale, NJ: Lawrence Erlbaum.
Selman, R. L. (2003). *The promotion of social awareness.* New York: Sage.
Seligman, M. (1995). *The Optimistic Child.* New York: Houghton Mifflin Co. (1995).
Singh, B. R. (1991). Teaching methods for reducing prejudice and enhancing academic achievement for all children. *Educational Studies, (2)*, 157-171.
Solomon, D., Watson, M.S., & Battistich, V.A., (2002). Teaching and school effects on moral/prosocial development. In V. Richardson (Ed.), *Handbook for Research on Teaching.* Washington, D.C.: American Educational Research Association.
Staub, E. (1978). *Positive social behavior and morality, Vol. 1: Social and personal influences.* New York: Academic Press.
Steyer, J.P. (2002). *The other parent* (Chapter 1, pp. 3-25). New York: Atria Books.
Timpson, W.M. (2002). *Teaching and learning peace.* Madison, WI: Atwood Publishing.
Wang, M.C., Haertel, G.D., & Walberg, H.J. (1998). Building Educational Resilience.*Phi Beta Kappa Fastbacks, 430*, 7-61.
Watson, M., & Eckert, L. (2003). *Learning to trust.* San Francisco: Jossey-Bass.
Zimmerman, B. J., Bonner, S., Kovach, R. (2002). *Developing self-regulated learners.* Washington, D.C.: American Psychological Association.

Darcia Narvaez, M. Div., Ph.D., is an Associate Professor at the University of Notre Dame. She has published extensively about moral development, moral education, and moral information processing.

Richard Herbst is a senior psychology and media theory major at the University of Notre Dame. He is the author of texts concerning social justice, child and migrant advocacy, and personal sanctuaries.

Scott Hagele is a senior at the University of Notre Dame, majoring in psychology and English. His research interests include the cognitive and developmental dimensions of character education.

Anna Gomberg, B.A., is a graduate student in developmental psychology at the University of Notre Dame. Her primary interests include moral development and media literacy.

Waging Peace Through Forgiveness Education in Belfast, Northern Ireland: A Review and Proposal for Mental Health Improvement of Children

Robert D. Enright
University of Wisconsin-Madison

Elizabeth A. Gassin
Olivet Nazarene University

Jeanette A. Knutson
University of Wisconsin-Madison

Violence and the threat of violence in Northern Ireland have been prevalent for centuries. We first review the literature showing the adverse psychological effects on children from terrorism and on-going community violence. Despite our knowledge of the negative psychological effects on children in the context of community violence, little research has actually been done on the amelioration of these effects. We propose here to introduce the concept of forgiveness into the Belfast schools as a way to counter children's anger, anxiety, and depression that can emerge when exposed to violence. After reviewing the existing empirical studies that validate forgiveness as a way to reduce negative psychological effects of injustice, we describe forgiveness curricula that we are currently implementing in first grade classrooms and will implement in grades two through high school over the years. Forgiveness education, because it targets anger and resentment that too often accompany violence, may be the missing piece to the peace puzzle.

Northern Ireland has not known peace for centuries. Forgiveness may be a key to the long-awaited peace and it is our goal in this article to outline a long-term peace plan for that region based on forgiveness education programs in schools, homes, and places of worship. The long-term objective of the current project is to improve the mental health of the youngest generation of Belfast, Northern Ireland (NI), targeting those reactions to interpersonal hurt (anger, anxiety, and depression) that contribute to less-than-optimal mental health and probably are also proximal causes of interpersonal violence. Children in NI represent a unique risk group, as they are exposed to poverty and both acute and chronic forms of violence (terrorism and increasing community violence, respectively), a situation that is not found in the USA. Not surprisingly, the need for emotional regulation with children in Belfast is high. For example, in one elementary school of 212 children in which we are working, 100 of them are currently being treated for anxiety and depressive disorders. Based on the research reviewed below and our own experience in Belfast, we believe that the severe and on-going injustices create anger and fear in the children. Over time, these continual negative emotions can give rise to psychological disorders in some of the children (see Enright & Fitzgibbons, 2000, for a review of research studies showing links between anger and psychological disorders and the links between forgiveness and emotional restoration).

We seek to meet our overall objective through the development, implementation, and evaluation of a manualized forgiveness intervention in the schools and the provision to parents and clergy, materials related to forgiveness for use at home and in the churches. For this article, we will focus only on the school aspect of the program.

Background of the Conflicts in Northern Ireland

What is the context of our work in NI? First, we are working with a population that is impoverished. A recent report called Northern Ireland "one of the poorest areas of the European Union" ("Country report...", 2000, p. 3; see also Cairns & Darby, 1998) and noted that 37% of NI children

live in poverty. As recently as June 2002, NI was deemed the poorest area of the UK, having the lowest weekly full-time earnings, the highest level of income support claimants, and the lowest rate of economic activity in general of the entire UK ("Measuring poverty...", 2002). Secondly, we are working with two generations (primary school children and their parents and teachers) who have been exposed to two forms of violence: *terrorism* (violent, often unpredictable strategies designed to kill innocent people with the goal of political intimidating and controlling a group or nation; Gurwitch, Sitterle, Young, & Pfefferbaum, 2002) and *on-going community violence* (chronic threat of intentional yet seemingly-random violence, usually occurring in a limited geographic area and less likely to be politically-driven than terrorism or war; Kupersmidt, Shahinfar, & Voegler-Lee, 2002).

Both terrorism and on-going community violence are a threat in NI today. While completed terrorist attacks are not as common now as they were in the 1970s and 80s, we note that one of the worst car bomb attacks in the history of The Troubles occurred just less than five years ago (August 1998 in Omagh), and large car bombs are still being placed (and fortunately discovered in time and defused) in Belfast, most recently in March and May 2003.[1] In the last decade, bombing incidents (including defusions) have risen from a low of two in 1995 to 187 reported incidents in 2002 (with 349 reported in 2001; Police Service of Northern Ireland [PSNI], 2003). Crime rates, as reported by PSNI, have recently increased.[2] Offenses against persons (e.g., murder, aggravated assault) and sexual offenses have risen by 22%, while kidnapping rates have risen 71%. Members of both Catholic and Protestant subcultures in NI have a tradition of parading during the summer months, events which have often led to violence. In the recent past, disorder at these parades rose from six incidents in 2000 to 29 in 2001 (PSNI, 2003). Clearly, both today's children and the generation responsible for guiding them have been and are continuing to be exposed to multiple risks (poverty and violence), and as far as violence is concerned, to both acute and chronic risk.

For the most part scholars reviewing research on the psychological effects of The Troubles have claimed that persons who were adults during the height of The Troubles (who are now elderly) have coped well with the political violence they experienced (Cairns & Wilson, 1989, 1993; Curran & Miller, 2001). (We note, however, that people directly exposed to violence were clearly at risk for PTSD [Loughrey, Curran, & Bell, 1993]). The impact on children, however, may not have been so benign.[3] Most early works on the effects of The Troubles on children (e.g., Fraser, 1974) suggested some children experienced psychosomatic and behavior problems (e.g., asthma, sleep problems, stuttering) that seemed to be linked to anxiety arising from exposure to violence; perhaps not surprisingly, the next generation of publications questioned these early conclusions. However, in a 1985 conference paper, McGrath and Wilson (cited in Cairns & Wilson [1993], who praise the paper for its methodological rigor) report elevated psychoticism scores on the Junior Eysenck Personality Questionnaire in NI children in comparison to English children, as well as a moderate, significant correlation in the NI sample between exposure to violence and scores on the Rutter Scale (measuring behavioral disturbance). While some have concluded that children have weathered the stress of The Troubles relatively well (e.g., Cairns & Wilson, 1993), in a recent review of the impact of The Troubles on psychiatric services in NI, Curran and Miller (2001) report that referrals of children did increase after major acts of violence.[4] These authors maintain that children's prolonged exposure to conflict and reminders (via media and/or police presence) of violence exacerbated posttraumatic stress symptomatology, mood disorders, behavioral problems, and academic struggles.

It is not surprising that children may be more at risk than adults for both immediate and later-onset stress reactions after exposure to acute or chronic violence. Children may lack the cognitive resources and emotional coping strategies to deal effectively with such events. For this and other developmental reasons, they may be more dependent on adults around them for support in the face of violence, but those same adults may be hampered in supporting a child by their own attempts at coping with the situation. In addition, trauma experienced at a young age may affect developing neurological systems, creating a neuropsychological vulnerability for which persons exposed to trauma for the first time as adults would not be at risk (Perry & Pollard, 1998).

Recent Research on the Impact of Terrorism

Further evidence on the effect of terrorism on children comes from recent work in areas other than NI.[5] For example, research on the effects of the Oklahoma City bombing demonstrates that even indirect exposure to terrorist attacks is related to persistent post-traumatic stress symptomatology. Children living in Oklahoma City who were exposed to a great amount of media about the bombing but did not personally know a victim reported comparable levels of post-traumatic stress symptomatology as did those who lost a parent or sibling in the blast (Pfefferbaum, Nixon, Krug et al., 1999; Pfefferbaum, Nixon, Tivis et al., 2001; Pfefferbaum, Nixon, Tucker et al., 1999); children living 100 miles outside of Oklahoma City who had only indirect exposure to the tragedy were more likely to display traumatic stress symptoms than those who were not exposed to the event even indirectly, even two years after the bombing (Pfefferbaum, Seale, McDonald et al., 2000). Because this is a relatively new field, we cannot predict the outcome in adulthood of children's exposure to terrorism. However, we must agree with researcher Robin Gurwitch, who claims that for such children, "the potential exists for serious long-term psychological sequelae" (Gurwitch et al., 2002, p. 342).

As the paragraphs above make clear, a threat of terrorism still exists in NI and should overt acts of terrorism continue, some (if not many) of today's children will be at risk for various psychological difficulties. In fact, just the threat of

terrorism may be a risk factor for today's young people, for as prolific NI researcher Cairns (Cairns & Wilson, 1989) noted, early research in the effects of violence suggested that the highest levels of referrals for psychological help were found in areas that were under threat of attack but not experiencing overt violence. (Similar results were reported by Thabet et al., [2002], who found that Palestinian children *indirectly* exposed to violence demonstrated higher levels of anxiety than those who were *directly* exposed.) The material presented above also suggests that the teachers and parents of today's young children in NI may be carrying some baggage from their own childhood and youth, which was dominated by horrific acts of terrorism. Adults' experiences filter down to the children of today. Recent empirical work demonstrates that a substantial minority of NI's young children absorb the anxieties, prejudices, and hurts of the older generation, especially after they begin schooling (Connolly, Smith, & Kelly, 2002). In addition, parents who directly experienced violence may pass on trauma-related vulnerability to their children (e.g., Harkness, 1993; Baranowsky, Young, Johnson-Douglas, Williams-Keller, McCarrey, 1998; Kellermann, 2001a, 2001b).[6] Subsequently, any intervention seeking to promote forgiveness in young children in NI must also include these adults, who themselves probably have serious interpersonal hurts in their pasts. Our intervention does that.

Recent Research on the Impact of Community Violence

The offspring of the terrorism of the 20th century is increasing acts of community violence, often based on ethnic affiliation and/or drug use. Recent qualitative and quantitative reviews of psychological correlates of exposure to community violence demonstrate that children who are exposed to – or in some cases only hear about - community violence are at risk for elevated levels of depression, anger, anxiety, sleep problems, externalizing behavior problems, and symptoms that parallel those of PTSD (Fletcher, 1996; Garbarino, Dubrow, Kostelny, & Pardo, 1992; Groves, 2002; Mazza & Overstreet, 2000; Osofsky, 1995; Kupersmidt, Shahinfar, & Voegler-Lee, 2002). Not surprisingly, academic problems are also common and may well be caused (at least in part) by the emotional consequences described above (Garbarino et al., 1992; Mazza & Overstreet, 2000; Schwartz & Gorman, 2003).

As is often the case with the impact of risk factors on children, the relationship between exposure to community violence and adverse psychological outcomes is moderated by other variables more proximal to the child. Research suggests that the main moderating variable here is the family: children from supportive families (operationalized in various ways in various studies) are less likely to show adverse effects of exposure to violence than those in unsupportive families (Gorman-Smith & Tolan, 1998; Kliewer, Lepore, Oskin, & Johnson, 1998; Martinez & Richters, 1993; Overstreet, Dempsey, Graham, & Moely, 1999; Richters & Martinez, 1993). These recent studies affirm what the resilience research (e.g., Garmezy, 1993; Werner & Smith, 1982) has been demonstrating for some time: family relationships are crucial to helping a child overcome risk factors such as exposure to community violence. This is probably especially true for young children (Fletcher, 1996; Green et al. 1991).

Although not widely discussed as a moderator of the impact of exposure to community violence on children, we note that school-related variables have received some attention as protective factors in the resiliency literature, both in qualitative reviews (e.g., Garbarino et al. 1992; Kirby & Fraser, 1997; Roeser, Eccles, & Strobel, 1998) and recent empirical studies (e.g., Anson, 1995; Arthur, Hawkins, Pollard, Catalano, & Barglioni, 2002; Botcheva, Feldman, & Leiderman, 2002). Because our intervention encourages classrooms to become forgiving communities, our program should help improve classroom climate, which in turn may help protect children against the adverse psychological effects of violence and terrorism in the larger community.

The Significance of Forgiveness Education for the Children of NI

We begin to see the following picture developing. Children in today's NI are being raised by people who themselves have directly experienced the pain of interpersonal violence in the form of terrorism. The threat of terrorism still hangs over the head of these children, and the experience of regular community violence is an unfortunate fact of their existence. The impact of the violence around these children is probably also exacerbated by the fact that many of them are experiencing risk accumulation, being exposed to violence *and* living in poverty (Cairns & Toner, 1993).

Anger, anxiety, and depression linked with interpersonal hurt, then, are relevant on at least two levels. First, these children risk "absorbing" these emotions from the adults in their environment, who themselves were raised during a very violent period of NI's history. In turn, on the foundation of this emotional inheritance, children may victimize or be victimized by others, thereby intensifying these feelings and their consequences. Secondly and relatedly, these children are at risk for being exposed to, if not actually involved in, overt acts of terrorism and/or community violence. As a result, these children may experience further increases in anger, anxiety, and/or depression, and their consequences (poorer mental health, academic performance, and so forth).

To our knowledge, no well-researched intervention in NI has demonstrated effectiveness in reducing anger, anxiety, and depression in children. Peace education efforts, which involve teaching people about peaceful conditions and giving them the skills for creating such conditions (Harris, 1988), do exist in NI (Coolahan, cited in Rogers, 1991; Duffy, 1992, 2000); however, they generally do not target these key emotions and are notoriously under-evaluated. This is true of what might be considered Northern Ireland's most comprehensive peace education program, Education for Mutual Understanding (EMU; Dunn & Morgan, 1999).

Integrated schools (where Catholic and Protestant children go to school together) also exist, but articles found through a search of standard databases (PsycINFO, ERIC, the internet) failed to demonstrate sound empirical evaluation of their effects on mental health, although there are some qualitative and survey studies that report a positive effect of this form of schooling on behavioral and attitudinal variables, such as relationships between Catholic and Protestant children and levels of tolerance (e.g., Cairns & Toner, 1993; Johnson, 2000; McGlynn, 2000; McWhirter, 1983). As will be seen below, 20 years of forgiveness research shows the efficacy of forgiveness education for improving mental health. Therefore, we claim with some degree of confidence that forgiveness education in the school and home from the early grades is both 1) a preventative measure to help reduce the negative effects of living in a dangerous environment (ultimately reducing the level of danger itself), and 2) an intervention to help children who are currently struggling with exposure to violence and the negative psychological consequences thereof.

In addition to helping the children of NI, the program piloted and assessed in this study should also prove useful for application in the USA and other countries, as issues of personal (e.g., bullying and other forms of interpersonal violence) and systemic (e.g. poverty, discrimination) injustice are prevalent in many areas.

Preliminary Studies on Interpersonal Forgiveness

The Concept of Forgiveness

Our research group at the University of Wisconsin-Madison has pioneered work in the psychology of forgiveness over the past 18 years. Conceptually, forgiveness is defined as follows:

> People, upon rationally determining that they have been unfairly treated, forgive when they willfully abandon resentment and related responses (to which they have a right), and endeavor to respond to the wrongdoer based on the moral principle of beneficence, which may include compassion, unconditional worth, generosity, and moral love (to which the wrongdoer, by nature of the hurtful act or acts, has no right) (Enright & Fitzgibbons, 2000, p. 29).

A definition more amenable to psychological study was also offered by Enright (Enright and the Human Development Study Group, 1991): from this perspective, forgiveness can be defined as overcoming negative thoughts, feelings, and behaviors directed at an offender, and developing positive thoughts, feelings, and behaviors vis-à-vis the same. Before beginning empirical work on the topic, our group thoroughly reviewed philosophical work on forgiveness, which makes clear that forgiveness is offered from a position of strength (Enright et al., 1991). Forgiveness does not make one weak or vulnerable; it should not be confused with condoning (e.g., ignoring or subtly approving) an offense or with reconciliation (re-establishing a relationship with an offender). Neither does it preclude *moderate, limited* expressions of anger or seeking *reasonable* redress of injustice. Because forgiveness is a specific personal response to injustice and because negative reactions to injustice appear to be at the root of developing unhealthy levels of anger, anxiety, and depression that are prevalent in children in Belfast, it follows that forgiveness intervention should be appropriate in this context.

Theory and research demonstrates that the process of forgiveness can be broken down into 20 units (see Table 1). A person need not progress through all units in order but may flow between units as he or she forgives. These can be generalized into the four broad phases of *uncovering* (admitting the fact of the offense and experiencing its negative consequences), *deciding* (feeling a need for change and deciding to forgive), *working* (trying to see the offender through different eyes and with a softened heart), and *deepening* (finding meaning and purpose in the offense and experiencing the benefits of forgiveness).

Experimental research on forgiveness education

Early forgiveness education programs based on our work were implemented with adults and adolescents. Freedman and Enright (1996) experimentally demonstrated that using a forgiveness education program in individual counseling with survivors of incest led to an increase in forgiveness of offender, self-esteem, and hope, as well as a decrease in depression and anxiety. A similarly-designed study found similar results with men whose partners had chosen to have an abortion (Coyle & Enright, 1997): participants in the experimental group showed significant greater gains in forgiveness and significant greater decreases in anxiety, anger, and grief than did control subjects. Al-Mabuk, Enright, and Cardis (1995) demonstrated that a forgiveness intervention conducted with groups of parentally love-deprived college students was effective in increasing forgiveness, hope, self-esteem, and positive attitudes toward parents, while significantly decreasing anxiety. Using a 12-session forgiveness intervention with a similar group (college students who demonstrated non-secure attachment to their mothers), Lin (1998) was able to help participants not only decrease levels of anger and anxiety, but also improve attachment to their mothers. This is important because, according to current understanding of psychologists (e.g., Bowlby, 1969/1982, 1973, 1980; Cassidy & Shaver, 1999), attachment begins developing in infancy and can be very difficult to change, especially when the participants are already adults and presumably have had such an attachment style for years. Furthermore, empirical work demonstrates that non-secure attachment patterns can have deleterious effects on psychological and social functioning (see Cassidy & Shaver, 1999, for an extensive

review), so any intervention that improves attachment status may have implications for improving mental and relational health.

The current generation of forgiveness interventions includes successful work with groups demonstrating relatively serious psychopathology; it also moves beyond forgiving offenses at the level of the individual to the level of the couple. Lin (2001) showed that substance abusers in a residential program who went through a forgiveness therapy program demonstrated significantly more forgiveness and self-esteem and significantly greater decreases in depression, anger, anxiety, and vulnerability to drug use than those who went through a different form of therapy. Chapman et al.'s (2001) forgiveness intervention with men who had committed crimes and were found not guilt by reason of mental disease promoted significantly more forgiveness, hope, and self-esteem in experimental participants than in control. Recent research from our program demonstrates that forgiveness interventions with married couples are at least as effective as therapy based on Beck's powerful cognitive-behavioral model in promoting forgiveness, marital adjustment, family cohesion, and self-esteem, as well as decreasing anger, anxiety, and depression (Knutson, 2003).

Forgiveness education is also effective in facing end-of-life issues. In what is perhaps the first experimental evaluation of forgiveness education, Hebl and Enright (1993) demonstrated the efficacy of forgiveness in a group setting with elderly women, who after participating in the program demonstrated significantly more forgiveness than controls. More recently, Hansen (2002) showed that implementing a forgiveness intervention with terminal cancer patients led to significantly greater increases in forgiveness of offender, hope, and quality of life, and also to significantly less anger than a control intervention.

Given the findings from our studies, it is not surprising that a recent meta-analysis (Baskin & Enright, in press) of forgiveness programs for adults and older adolescents shows that forgiveness interventions implemented with *groups* led to an average increase of .83 of a standard deviation in forgiveness of offender and .59 of a standard deviation in mental health of offended party. Interventions used with *individuals* led to an average increase of 1.66 standard deviations for forgiveness and 1.42 standard deviations for mental health of offended party. All of these effect sizes can be characterized as "large" (Lipsey, 1990), supporting the claim that forgiveness interventions are powerful in improving the lives of those who participate in them.[7] It is also important to note that most of the studies from our research group demonstrated that gains in forgiveness and mental health are maintained over time, even up to a year later (e.g., Freedman & Enright, 1996).

What about the impact of forgiveness education with children? An initial effort (Hepp-Dax, 1996) applying our model to children led to mixed results. The researcher did a four-week forgiveness intervention with inner-city 5th graders and found that although the experimental group was significantly higher than an alternative-treatment control group in forgiveness at post-test, the control group did not replicate these findings when it was given the intervention. Seven months after the end of the intervention with the control-turned-intervention group, all children (both initial experimental and controls) showed improvements in forgiveness and self-esteem. However, this last finding is clearly quasi-experimental, so interpreting the findings is difficult. A four-week intervention may be too brief for children.

More recent work clearly suggests forgiveness education with children is effective in promoting psychological and social health. Especially germane here is Gambaro's (2002) work with young adolescents who demonstrated higher-than-average levels of anger. Gambaro implemented a forgiveness intervention with especially angry adolescents (age 12-14) in a group setting, while offering students in a control condition a Rogerian-based support group experience. Over the course of this twelve-week program, experimental group participants read two fictional stories about people who had suffered serious offenses at the hands of another. During any given session, a section of each story illustrating a particular part of the forgiveness process (see Table 1) was read, and students discussed the events in their story and in their own story of offense (which had been identified during pre-test with the Enright Forgiveness Inventory for Children. Part of each session was also dedicated to discussing current events at school and in the home and how forgiveness might be practiced in these situations. At post-test, participants in the experimental group demonstrated greater decreases in "anger as a trait," "having an angry temperament," and "predilection to react in an angry manner" than did control participants. In addition, those in the experimental group showed significantly greater improvements in attitudes towards school and family and in quality of relationships with friends and family. These results were maintained at follow-up nine months after the intervention ended. At follow-up Gambaro also looked at school grades, detentions, and in-school suspensions and found that the experimental group was significantly higher in academic achievement and significantly lower in detentions and suspensions than the control group.

Park (2003) continued Gambaro's emphasis on angry children, assessing the impact of a forgiveness intervention on female adolescents in Korea who were aggressive victims of peer abuse. In other words, they had been victimized by a peer and in turn became aggressors themselves. Park randomly assigned adolescents identified by a screening instrument to one of three groups: a 12-week forgiveness intervention (based on a culturally- and developmentally-adapted version of Enright's [2001] *Forgiveness is a Choice*), a 12-week "skill-streaming" program (enhancing prosocial skill development), or a no-contact control. Park's experimental forgiveness program is unique in that it took young people through the process of offering forgiveness and introduced them to seeking and receiving forgiveness as well. The first eight sessions were dedicated to an overview

Table 1

The phases and units of forgiveness (Enright & Fitzgibbons, 2000, p. 68)

UNCOVERING PHASE

1. Examination of psychological defenses
2. Confrontation of anger; the point being to release, not harbor, the anger
3. Admittance of shame, when this is appropriate
4. Awareness of depleted emotional energy
5. Awareness of cognitive rehearsal
6. Insight that the injured party may be comparing self with the injurer
7. Realization that oneself may be permanently and adversely changed by the injury
8. Insight into a possibly altered "just world" view

DECISION PHASE

9. A change of heart/conversion/new insights that old resolution strategies are not working
10. Willingness to consider forgiveness as an option
11. Commitment to forgive offender

WORK PHASE

12. Reframing, through role-taking, who the wrongdoer is by viewing him or her in context
13. Empathy and compassion toward the offender
14. Bearing/accepting the pain
15. Giving a moral gift to the offender

DEEPENING PHASE

16. Finding meaning for self and others in the suffering and in the forgiveness process
17. Realizing that self has needed others' forgiveness in the past
18. Insight that one is not alone
19. Realization that one may have a new purpose in life because of the injury
20. Awareness of decreased negative affect and, perhaps, increased positive affect, if this begins to emerge toward the injurer; awareness of internal, emotional release

Note: This table is an extension of Enright et al. (1991). Details of each unit can be found in Enright (2001).

of the model of offering forgiveness (as seen in Table 1), while the last four led participants on an exploration of seeking and receiving forgiveness for their own offenses. While at no time were students forced to offer or seek forgiveness, the last lesson did require students to develop an action plan based on the information and activities they encountered during the intervention. At post-test, the group who participated in the forgiveness intervention demonstrated greater increases in forgiveness and greater decreases in anger, self-reported delinquency, self-reported aggression, and hostile attributions than did the other groups, which did not differ from one another. These gains were maintained at the six-week follow-up. In addition, by follow-up the forgiveness group demonstrated more empathy

than the other groups.

We consider these latter studies, with their emphasis on reduction of anger and aggression in particular, critical for designing an effective program to help individuals heal from interpersonal violence and to quell The Troubles in Northern Ireland. In their review of the issue, Enright and Fitzgibbons (2000) established the decisive role of anger in the development of psychological and relational problems. Positive anger management is also critical for Northern Ireland. In essence, The Troubles are a series of anger- and revenge-promoting interpersonal offenses, originally based on religio-ethnic group membership, but more recently becoming based on gang affiliation and drug use patterns. The above literature demonstrates that forgiveness is effective in reducing anger, and yet it encourages people to actively work for positive change in personal relationships and society at large. The current forgiveness education intervention should reduce this toxic anger in children (and in those adults closest to them).

While anger may be the most obvious link between interpersonal hurt and both personal psychological dysfunction and the interpersonal violence that exists in NI, we also note that the empirically-demonstrated benefits of forgiveness in terms of depression and anxiety will also help individuals and the Belfast community heal.

A forgiveness treatment manual for counselors and laypersons based on this research is now published, *Forgiveness Is a Choice* (Enright, 2001); psychiatrists and psychologists can also consult *Helping Clients Forgive: An Empirical Guide for Resolving Anger and Restoring Hope* (Enright & Fitzgibbons, 2000) for a detailed discussion of more scholarly and professional issues involved in implementing forgiveness interventions. A children's picture book is in production, *Rising Above the Storm Clouds* (Enright, in press).

Intervention Plan for the Schools of Northern Ireland

In this section, we describe the general content of our school-based intervention with students in Belfast. Our first year within 12 first grade (Primary 3 in NI) classrooms is now completed and is awaiting analysis of the pretest to post-test evaluation data.

Throughout our manualized intervention for primary school children, we are targeting a particular aspect of children's social-cognitive development: reframing (Unit 12 in Table 1), in which the child understands that all people, even those who are unfair, have worth. Through our manuals, children are and will be taught about inherent worth of all people and to act on this insight by displaying the moral qualities of moral love (acting more out of a concern for the well-being of another than for oneself), kindness, respect, and/or generosity to those around them, including the ones who have hurt them. These five elements (reframing for inherent worth, moral love, kindness, respect, and generosity) are the focus of the first grade program. In the program we are careful to emphasize the distinction between forgiveness and reconciliation. A child does not reconcile with an unrepentant bully, for example.

These elements are key in the work phase of the model of forgiveness presented in Table 1, which has been the crux of all forgiveness interventions conducted to date.

The forgiveness education manuals planned for first grade through high school are developmental, with increasing cognitive complexity across the years. In the first grade manual, we introduce the children to the concept of forgiveness with the five themes above. To emphasize these themes, we have a three-part curriculum. The first part simply introduces these five concepts outside the context of forgiveness through the story-medium of Dr. Seuss's books. Part 2 introduces these five concepts again, but this time *within the context of forgiveness*, using stories again by Dr. Seuss. Part 3 introduces these five concepts within the context of the *child's own attempt to forgive someone.* In the second grade manual, we will repeat these foundational themes through Disney stories and DVDs. The third grade manual will go into greater depth in linking the moral principle of beneficence to forgiveness, which will be illustrated through more advanced literature such as *The Velveteen Rabbit* (Bianco, 1987) and Rising Above the Storm Clouds (Enright, in press). The outlines for the fourth grade manual and beyond are still being developed, and therefore are not discussed here.

Throughout the curriculum, the teachers make the important distinction between *learning* about forgiveness and *choosing to practice* it in certain contexts. Children are always free to try or not try forgiveness in response to their own personal hurts borne out of unfair treatment. In our experience in Belfast to date, children willingly try forgiveness when they are free to choose the person who was unfair to them and the event that each child considers to be unfair. The child's own classroom teacher will deliver the curriculum, to insure cultural and religious sensitivity regarding the nuances of forgiveness. Each year, we will evaluate the effectiveness of the programs with psychometrically-sound instruments given to children and teachers as well as with direct observation children's interactions.

Expectations and Hopes

Children who participate in forgiveness education in the school, home, and place of worship should show greater emotional regulation (less anger, anxiety, and depression), less anti-social behavior and more pro-social behavior than those who have not participated in the program; these gains should be maintained at follow-up one year later. We especially expect these outcomes for children who are identified as being "at-risk" for anger, anxiety, and/or depression, based on our previous findings.

We do not yet know if a comprehensive, long-term forgiveness intervention can positively affect the mental health of children (and their teachers and parents) living in

poverty and violence, although based on past research we expect it will. Demonstrating this may verify the usefulness of forgiveness education for promoting mental health (and the characteristics that go with it: productiveness, longer life, etc.) even in the most vulnerable areas of society.

In the long-run, our hope is that the children will grow into adulthood with the tools of forgiveness, given their learning over many years. Thus equipped, these children, now adults, may be less quick to act on anger and more quick to forgive before trying to implement justice. As the history of The Troubles has too often confirmed, attempts at justice when so many in the community are angry do not lead to the most satisfying of outcomes. Forgiveness may temper this and help promote greater fairness throughout society.

References

Al-Mabuk, R. H., Enright, R. D., & Cardis, P. (1995). Forgiveness education with parentally love-deprived late adolescents. *Journal of Moral Education, 24*, 427-443.

Anson, A. R. (1995). Risk and protection during the middle school transition: The role of school climate in early adolescent development. *Dissertation Abstracts International – B, 56*(03), 1716.

Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Barglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities that Care Youth Survey. *Evaluation Review, 26*, 575-601.

Baranowsky, A. B., Young, M., Johnson-Douglas, S., Williams-Keller, L., McCarrey, M. (1998). PTSD transmission: A review of secondary traumatization in Holocaust survivor families. *Canadian Psychology, 39*, 247-256.

Barber, B. K. (1999). Political violence, family relations, and Palestinian youth functioning. *Journal of Adolescent Research, 14*, 206-230.

Baskin, T. W., & Enright, R. D. (in press). Intervention studies of forgiveness: A meta-analysis. *Journal of Counseling and Development.*

Bianco, M. W. (1987). *The velveteen rabbit.* San Diego, CA: Harcourt, Brace, Jovanovich.

Botcheva, L. B., Feldman, S. S., & Leiderman, P. H. (2002). Can stability in school processes offset the negative effects of socio-political upheaval on adolescent adaptation? *Youth and Society, 34*, 55-88.

Bowlby, J. (1969/1982). Attachment and loss: Vol. 1. *Attachment.* New York: Basic Books.

Bowlby, J. (1973). Attachment and loss: Vol. 2. *Separation.* New York: Basic Books.

Bowlby J. (1980). Attachment and loss: Vol. 3. *Loss.* New York: Basic Books.

Cairns, E., & Darby, J. (1998). The conflict in Northern Ireland: Causes, consequences, and controls. *American Psychologist, 53*, 754-760.

Cairns, E., & Toner, I. J. (1993). Children and political violence in Northern Ireland: From riots to reconciliation. In L. A. Leavitt & N. A. Fox (Eds.), *The psychological effects of war and violence on children* (pp. 215-229). Hillsdale, NJ: Lawrence Erlbaum.

Cairns, E., & Wilson, R. (1989). Mental health aspects of political violence in Northern Ireland. *International Journal of Mental Health, 18*, 38-56.

Cairns, E., & Wilson, R. (1993). Stress, coping, and political violence in Northern Ireland. In J. P. Wilson & B. Raphael (Eds.), *International Handbook of Traumatic Stress Syndromes* (pp. 365-376). New York: Plenum.

Cassidy, J., & Shaver, P. R. (Eds.). (1999). *Handbook of attachment: Theory, research, and clinical applications.* New York: Guilford Press.

Chapman, R. F., Maier, G., Owen, A., Nousse, V., Park, J. H., & Enright, R. D. (2001, July). *Healing forgiveness; Group cognitive therapy for abused male forensic patients.* Paper presented at the World Congress of Behavioral and Cognitive Therapies, Vancouver, Canada.

Connolly, P., Smith, A., & Kelly, B. (2002). *Too young to notice? The cultural and political awareness of 3-6 year olds in Northern Ireland.* Belfast: Community Relations Council.

County report: *Northern Ireland* (2000). Belfast: Save the Children UK. Retrieved July 7, 2003, from http://www.savethechildren.org.uk/development/reg_pub/country_reports/NIreland_2000.pdf

Coyle, C. T., & Enright, R. D. (1997). Forgiveness intervention with postabortion men. *Journal of Counseling and Clinical Psychology, 65*, 1042-1046.

Curran, P. S., & Miller, P. W. (2001). Psychiatric implications of chronic civilian strife or war: Northern Ireland. *Advances in Psychiatric Treatment, 7*, 73-80.

Desivilya, H. S., Gal, R., & Ayalon, O. (1996). Extent of victimization, traumatic stress symptoms, and adjustment of terrorist assault survivors: A long-term follow-up. *Journal of Traumatic Stress, 9*, 881-889.

Duffy, T. (1992). *Peace education in a hostile environment: The divided society of Northern Ireland.* (ERIC Document Reproduction Service No. 358066)

Duffy, T. (2000). Peace education in a divided society: Creating a culture of peace in Northern Ireland. Prospects: *Quarterly Review of Comparative Education, 30*, 15-29.

Dunn, S., & Morgan, V. (1999). 'A fraught path' – education as a basis for developing improved community relations in Northern Ireland. *Oxford Review of Education, 25*, 141-153.

Enright, R. D. (2001). *Forgiveness is a choice: A step-by-step process for resolving anger and restoring hope.* Washington, DC: APA Life Tools.

Enright, R. D. (in press). *Rising above the storm clouds.* Washington, DC: Magination Press (American Psychological Association).

Enright, R. D., & Fitzgibbons, R. P. (2000). *Helping clients forgive: An empirical guide for resolving anger and*

restoring hope. Washington, DC: American Psychological Association.

Enright, R. D., & the Human Development Study Group. (1991). The moral development of forgiveness. In W. Kurtines & J. Gewirtz (Eds.)., *Handbook of moral behavior and development, Vol. 1* (pp. 123-152). Hillsdale, NJ: Lawrence Erlbaum.

Enright, R. D., & the Human Development Study Group. (1994). Piaget on the moral development of forgiveness: Identity or reciprocity? *Human Development, 37*, 63-80.

Fletcher, K. E. (1996). Childhood posttraumatic stress disorder. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (pp. 242-276). New York: Guilford.

Fraser, M. (1974). *Children in conflict*. Middlesex, England: Penguin Books.

Freedman, S. R., & Enright, R. D. (1996). Forgiveness as an intervention goal with incest survivors. *Journal of Counseling and Clinical Psychology, 64*, 983-992.

Gambaro, M. E. (2002). *School-based forgiveness education in the management of trait anger in early adolescents*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Garbarino, J., Dubrow, N., Kostelny, K., & Pardo, C. (1992). *Children in danger: Coping with the consequences of community violence*. San Francisco, CA: Jossey-Bass.

Garbarino, J., & Kostelny, K. (1996). The effects of political violence on Palestinian children's behavior problems: A risk accumulation model. *Child Development, 67*, 33-45.

Garmezy, N. (1993). Children in poverty: Resilience despite risk. *Psychiatry, 56*, 127-136.

Gorman-Smith, D., & Tolan, P. (1998). The role of exposure to community violence and developmental problems among inner-city youth. *Development and Psychopathology, 10*, 101-116.

Green, B. L., Korol, M., Grace, M. C., Vary, M. G., Leonard, A. C., Gleser, G. C., et al. (1991). Children and disaster: Age, gender, and parental effects on PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 945-951.

Groves, B. M. (2002). *Children who see too much: Lessons from the Child Witness to Violence Project*. Boston: Beacon Press.

Gurwitch, R. H., Sitterle, K. A., Young, B. H., Pfefferbaum, B. (2002). The aftermath of terrorism. In A. M. LaGreca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 327-358). Washington, DC: American Psychological Association.

Hansen, M. J. (2002). *Forgiveness as an educational intervention goal for persons at the end of life*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Harkness, L. L. (1993). Transgenerational transmission of war-related trauma. In J. P. Wilson & B. Raphael (Eds.), *International Handbook of Traumatic Stress Syndromes* (pp. 635-643). New York: Plenum.

Harris, I. M. (1988). *Peace education*. Jefferson, NC: McFarland & Co.

Hebl, J. H., & Enright, R. D. (1993). Forgiveness as a psychotherapeutic goal with elderly females. *Psychotherapy, 30*, 658-667.

Hepp-Dax, S. H. (1996). *Forgiveness as a goal with fifth-grade inner-city children*. Unpublished doctoral dissertation, Fordham University, New York.

Hoge, C., Pavlin, J. A., & Milliken, C. S. (2002). Psychological sequelae of September 11. *New England Journal of Medicine, 347*, 443.

Johnson, L. (2000). *The impact of integrated education on students' tolerance of diversity: An ethnographic case study*. Unpublished manuscript, Hofstra University.

Kellermann, N. P. F. (2001a). The long-term psychological effects and treatment of Holocaust trauma. *Journal of Loss and Trauma, 6*, 197-218.

Kellermann, N. P. F. (2001b). Psychopathology in children of Holocaust survivors: A review of the research literature. *Israel Journal of Psychiatry, 38*, 36-46.

Kirby, L. D., & Fraser, M. W. (1997). Risk and resilience in childhood. In M. W. Fraser (Ed.), *Risk and resilience in childhood: An ecological perspective* (pp. 10-33). Washington, DC: National Association of Social Workers.

Kliewer, W., Lepore, S. J., Oskin, D., & Johnson, P. D. (1998). The role of social and cognitive processes in children's adjustment to community violence. *Journal of Consulting and Clinical Psychology, 66*, 199-209.

Knutson, J. A. (2003). *Strengthening marriages through the practice of forgiveness*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Kupersmidt, J. B., Shahinfar, A., & Voegler-Lee, M. E. (2002). Children's exposure to community violence. In A. M. LaGreca, W. K. Silverman, E. M. Vernberg, & M. C. Roberts (Eds.), *Helping children cope with disasters and terrorism* (pp. 381-401). Washington, DC: American Psychological Association.

Lin, W. (2001). *Forgiveness as an educational intervention goal within a drug rehabilitation center*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Lin, W. N. (1998). *Forgiveness as an intervention goal with late adolescents with insecure attachment in Taiwan*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage Publications, Inc.

Loughrey, G. C., Curran, P. S., & Bell, P. (1993). Posttraumatic stress disorder and civil violence in Northern Ireland. In J. P. Wilson & B. Raphael (Eds.), *International Handbook of Traumatic Stress Syndromes* (pp. 377-383). New York: Plenum.

Martinez, P., & Richters, J. E. (1993). The NIMH Community Violence Project II: Children's distress symptoms

associated with violence exposure. *Psychiatry, 56,* 22-35.

Mazza, J. J., & Overstreet, S. (2000). Children and adolescents exposed to community violence: A mental health perspective for school psychologists. *School Psychology Review, 29,* 86-101.

McGlynn, C. (2000, April). *A case study of the impact of postprimary integrated education on past pupils in Northern Ireland.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

McWhirter, L. (1983). Contact and conflict: The question of integrated education. *Irish Journal of Psychology, 6,* 13-27.

Measuring poverty and social exclusion in Northern Ireland. (2002, June). Belfast: Northern Ireland Assembly, Research and Library Service.

Osofsky, J. D. (1995). The effects of exposure to violence on young children. *American Psychologist, 50,* 782-788.

Overstreet, S., Dempsey, M., Graham, D., & Moely, B. (1999). Availability of family support as a moderator of exposure to community violence. *Journal of Consulting and Clinical Psychology, 28,* 151-159.

Park, J. H. (2003). *Validating a forgiveness education program for adolescent female aggressive victims in Korea.* Unpublished doctoral dissertation, University of Wisconsin-Madison.

Perry, B. D., & Pollard, R. (1998). Homeostasis, stress, trauma, and adaptation: A neurodevelopmental view of childhood trauma. *Child and Adolescent Psychiatric Clinics of North America, 7,* 33-51.

Pfefferbaum, B., Nixon, S. J., Krug, R. S., Tivis, R. D., Moore, V. L., Brown, J. M., et al. (1999).. Clinical needs assessment of middle and high school students following the 1995 Oklahoma City bombing. *American Journal of Psychiatry, 156,* 1069-1074.

Pfefferbaum, B., Nixon, S. J., Tivis, R. D., Doughty, D. E., Pynoos, R. S., Gurwitch, R. H., et al. (2001). Television exposure in children after a terrorist incident. *Psychiatry, 64,* 202-211.

Pfefferbaum, B., Nixon, S. J., Tucker, P. M., Tivis, R. D., Moore, V. L., Gurwitch, R. H., et al. (1999). Posttraumatic stress responses in bereaved children after the Oklahoma City bombing. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1372-1379.

Pfefferbaum, B., Seale, T. W., McDonald, N. B., Brandt, E. N., Rainwater, S. M., Maynard, B. T. et al. (2000). Posttraumatic stress two years after the Oklahoma City bombing in youths geographically distant from the explosion. *Psychiatry, 63,* 358-370.

Police Service of Northern Ireland [PSNI]. (2002). Crime statistics (various files). Retrieved May 6, 2003 from http://www.psni.police.uk/stats/index.shtml.

Richters, J. E., & Martinez, P. E. (1993). Violent communities, family choices, and children's chances: An algorithm for improving the odds. *Development and Psychopathology, 5,* 609-627.

Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health: Selected issues and empirical illustrations at the level of the individual. *Educational Psychologist, 33,* 153-176.

Rogers, P. (1991). *Education for peace in the classroom-curriculum development, strategies, and materials: A case study from Ireland.* ERIC Document 369 693

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliot, M. N., et al. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine, 345,* 1507-1512.

Schwartz, D., & Gorman, A. H. (2003). Community violence exposure and children's academic functioning. *Journal of Educational Psychology, 95,* 163-173.

Thabet, A. A. M., Abed, Y., & Vostanis, P. (2002). Emotional problems in Palestinian children living in a war zone: A cross-sectional analysis. *Lancet, 359,* 1801-1804.

Trautman, R., Tucker, P., Pfefferbaum, B., Lensgraf, S. J., Doughty, D. E., Buksh, A., et al. (2002). Effects of prior trauma and age on posttraumatic stress symptoms in Asian and Middle Eastern immigrants after terrorism in the community. *Community Mental Health Journal, 38,* 459-474.

Waltman, M. A. (2002). *The psychological and physiological effects of forgiveness education in male patients with coronary artery disease.* Unpublished doctoral dissertation, University of Wisconsin-Madison.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth.* New York: Adams-Bannister-Cox.

[2] PSNI notes that some of the increase may be due to improved reporting procedures, but these improved procedures mostly concern "low-level" crimes. What is meant by "low-level" is not clear, but it is unlikely that it refers to the obvious offenses mentioned here.

[3] On a related note, Kellermann (2001a) has clinically noted an inverse relationship between age during the Holocaust and degree of psychological impact of the experience in child survivors of concentration camps; similar evidence based on quantitative research can be found in Fletcher (1996) and Trautman et al. (2002).

[4] This parallels reports of child referrals in the US after the terrorist attacks of September 11, 2001 (Hoge, Pavlin, &

Footnotes

[1] Several of the recent bomb threats seem tied to US politics. The March 2003 bomb was placed and defused right before a meeting between UK officials and President George W. Bush, while a bomb placed and defused in December 2000 occurred during then US President Bill Clinton's visit to Belfast.

Milliken, 2002; see also Schuster et al., 2001).

[5] Due to space considerations and to provide an example of how exposure to one self-contained act of terrorism affects children, here we have decided to concentrate on the impact of the Oklahoma City bombing. Another relevant situation (which in some ways parallels the situation in NI) is the recent history between Israelis and Palestinians. Suffice it to say that literature on the impact of terrorism and other forms of violence in that geographical area show that children of both communities who have been either directly or indirectly exposed to violence demonstrate emotional and behavioral problems as a result (e.g., Barber, 1999; Desivilya, Gal, & Ayalon, 1996; Garbarino & Kostelny, 1996; Thabet, Abed, & Vostanis, 2002).

[6] Data from 1985 (McGrath & Wilson, cited in Cairns & Wilson, 1993) show that 20% of 10-11 year olds at that time reported being in or witnessing a car bombing. McGrath and Wilson's young people are probably about the same age as the parents of today's Primary 3 (first grade) pupils. Loughrey, Curran, and Bell (1993) report that 23% of people exposed to violent acts in NI develop PTSD, a number comparable to those who are exposed to other forms of trauma. This leaves a very conservative estimate of about 5% of young adults today who have a history of PTSD. When one considers that the car bombs about which McGrath and Wilson's young subjects reported are only one form of violence in NI (assassinations, violent riots, and other forms of brutality are also prevalent), and that most children have more than one significant adult in their lives, it becomes moderately probable that any given child today has at least one significant adult in their lives with a history of PTSD related to The Troubles.

[7] Aside from "typical" mental health measures, forgiveness education is beginning to show promise for improving physical health as well. Waltman (2002) demonstrated experimentally that male cardiac patients who had been through a forgiveness program had better cardio-vascular health than those in a control group when measured 10 weeks after participation in the program.

Robert Enright has been a professor of Educational Psychology at the University of Wisconsin-Madison for 25 years. His research group pioneered the scientific study of forgiveness in the mid-1980s.

Elizabeth A. Gassin spent three years teaching in various institutions of higher education in Russia, including serving one year as a Fulbright scholar. She now serves as Associate Professor of Psychology at Olivet Nazarene University at Bourbonnais, IL. Her research interests include the psychology of forgiveness and cultural and philosophical variations in the understanding of the concept.

Jeanette Knutson received her Ph.D. from the Educational Psychology Department at the University of Wisconsin-Madison. She is developing forgiveness education and conducting research in Northern Ireland.

Author Note

Inquiries about this article should be sent to the first author, Department of Educational Psychology, 1025 West Johnson Street, University of Wisconsin-Madison, Madison, WI 53706. email: renright@wisc.edu.

Effects of Perspective Reversal Training and Conflict-Resolution-Based Classroom Management in Kindergarten

Laurie Stevahn
Seattle University

Katie Oberle
Edina Public Schools

The effectiveness of a conflict resolution program was examined in a suburban K-5 public elementary school in the Midwestern United States. Participants were 38 children in two kindergarten classes. The two independent variables were (a) the presence versus the absence of perspective reversal training incorporated into curriculum-integrated conflict resolution training and (b) the presence versus the absence of a conflict-resolution-based classroom management strategy to resolve actual disputes among children. The issues examined included (1) the effectiveness of the conflict resolution training, (2) the impact of the perspective reversal training and the conflict-resolution-based classroom management strategy on the effectiveness of the conflict training, (3) the impact of the conflict training on cognitive-developmental growth in social understanding and interpersonal functioning, and (4) the impact of the conflict training on academic achievement.

Introduction

Although 5- and 6-year-olds in early childhood education settings frequently engage in win-lose conflicts (see Shantz, 1987), young children rarely are included in peace education initiatives that teach constructive conflict resolution procedures. Perhaps this omission occurs because many question the developmental capacity of primary-age children to learn how to manage conflict constructively. Doing so requires considering the interests of others as well as self-interests. Yet every major theory that addresses issues in social and cognitive development suggests that young children are not able to engage in social perspective taking (e.g., see Kohlberg, 1987; Piaget, 1928, 1932; Selman, 1980, 1981). Numerous developmental studies conducted in naturalistic contexts indicate that young children tend to engage in various forms of distributive negotiation, striving to maximize personal outcomes, often at the expense of others. In contrast, recent conflict training research indicates that kindergartners can learn to engage in integrative negotiation, which entails problem solving to find solutions that mutually satisfy everyone's interests and maximize joint outcomes (Stevahn, Johnson, Johnson, Oberle, & Wahl, 2000). Little information exists, however, on the impact of teaching young children conflict resolution procedures in school settings. The first purpose of this study was to examine the extent to which kindergartners can learn what conflict is and how to use integrative negotiation to resolve it.

The second purpose of this study was to examine what mediates the effectiveness of school-based conflict resolution training. Previous research indicates that conflict training in a cooperative classroom context is more effective than training in an individualistic context (Stevahn, Johnson, Johnson, & Real, 1996). Within a cooperative context, however, two additional factors may further enhance conflict training in schools: (1) the extent to which students learn and engage in perspective reversal during negotiations and (2) the routine use of integrative negotiation in classroom management. Perspective reversal essentially entails presenting the viewpoints and feelings of others (Johnson & Johnson, 2003). Previous studies primarily involving adults in college settings indicate that engaging in perspective reversal reduces common barriers and increases mutual understanding in conflict communications (Johnson, 1971). Furthermore, routinely using integrative negotiation in classrooms to manage actual disputes may enhance conflict training by involving students in sustained, meaningful, repeated practice that genuinely "counts" – thereby strengthening procedural learning. This study implemented both a perspective reversal treatment as well as a "real use" treatment to determine their respective effects on conflict

training.

The third purpose of this study was to obtain behavioral-observational data on the effectiveness of the conflict resolution training. Previous research on school-based conflict programs has relied heavily on questionnaires and interviews (Johnson & Johnson, 1996). The ultimate goal of conflict training, however, is for students to willingly and skillfully transfer their learning to real conflict situations. Although a few studies provide some behavioral evidence that trained children can and do use conflict procedures to resolve actual disputes (Johnson & Johnson, 2002), many more observations are needed to better understand the complexities of using conflict procedures in real situations. This study used a new behavioral measure that documented how all of the children dealt with an actual conflict over scarce resources.

The fourth purpose of this study was to determine whether conflict resolution training provides young children with a cognitive-developmental advantage in managing conflicts. Resolving conflict constructively requires recognizing that a conflict exists and attending to the interests of everyone involved. Untrained children tend not to exhibit those capabilities. Selman's (1980, 1981) model for classifying interpersonal negotiation strategies, for example, posits that children progress through four broad stages of development as they advance in age. Young children tend to deal with conflict at an undifferentiated/egocentric level of interaction characterized by impulsive "fight" or "flight" behavior, or at a differentiated/subjective level characterized by unilateral actions to "control" or "appease." It is not until children are beyond the age of 8 that they exhibit a self-reflective/reciprocal level of interaction characterized by "fair trades and exchanges," and not until children are beyond the age of 14 that they exhibit the most developmentally advanced level characterized by social perspective coordination that enables constructing mutually satisfying agreements. Although Stevahn et al. (2000) provided evidence that kindergartners both can learn and use integrative negotiation to resolve conflict situations, developmental measures of cognitive-social growth were not used to assess any developmental advantages that may have resulted. This is the first study to use such a measure.

The fifth purpose of this study was to determine the impact of teaching conflict resolution as an integrated component of the academic curriculum. Whether conflict programs are institutionalized in schools may depend on connections to the academic curriculum and effects on student achievement. Conflict training that is integrated into existing curricula may be more feasible to implement than training that remains separate and apart from daily curricular requirements; it also is less likely to further stress an already "overcrowded curriculum." Moreover, innovations in schools have tended to disappear unless they have had positive effects on student achievement (Fullan, 2001; Johnson, 1970). Although a growing number of curriculum-integrated conflict training studies show academic gains in achievement across a range of content areas and grade levels (Johnson & Johnson, 2002), additional data would further contribute to generalizing results. This study provides additional data by integrating conflict training into a required kindergarten thematic instructional unit on friendship.

Method

Participants

Participants were 38 kindergartners in two classes in a suburban K-5 public elementary school in Edina, Minnesota. The morning class consisted of 18 children (8 boys and 10 girls) who were randomly assigned to a curriculum-integrated-conflict-training-only condition (4 boys and 5 girls) or a curriculum-integrated-conflict-training plus perspective-reversal-training condition (4 boys and 5 girls). The afternoon class consisted of 20 children (10 boys and 10 girls) who were randomly assigned to a curriculum-integrated-conflict-training plus conflict-resolution-based-classroom-management condition (5 boys and 5 girls) or a curriculum-integrated-conflict-training plus perspective-reversal-training plus conflict-resolution-based-classroom-management condition (5 boys and 5 girls). The randomization resulted in each condition containing equivalent mixes of children by age (5- and 6-year-olds), gender (boys and girls), academic ability (diverse literacy and numeric skills), social-emotional development (various levels of maturity), and special needs.

Independent Variables

Two independent variables were included in this study: (a) the presence versus the absence of perspective reversal training incorporated into conflict resolution taught as an integrated component of a kindergarten curriculum unit on friendship and (b) the presence versus the absence of a conflict-resolution-based classroom management strategy used to resolve actual conflicts among children. This 2 x 2 design established four conditions.

Children in all conditions received conflict resolution training, which consisted of teaching a definition of conflict and how to negotiate to reach an integrative agreement. The six-step integrative negotiation procedure was (1) recognizing that a conflict exists and expressing a desire to resolve it constructively ("Stop. We have a conflict. Let's work it out."), (2) stating wants and underlying reasons ("I want ... because ..."), (3) expressing feelings and underlying reasons ("I feel ... because ..."), (4) communicating understanding of the other person ("I heard you say that you want ... and feel ... because ..."), (5) inventing three or more optional solutions that maximize mutual gain ("Some ideas are ..."), and (6) reaching agreement by jointly selecting and shaking hands on one of the options ("We agree on ..."). The entire procedure was taught as an integrated component of a kindergarten curriculum unit on friendship

(for sample activities see Stevahn et al., 2000). The conflict training occurred 30 minutes daily over 13 consecutive school days for a total of 6.5 hours of instruction.

Children in the perspective-reversal-training conditions also were taught to recognize a range of emotions that people commonly experience in conflicts (such as sad, frustrated, angry) and how to engage in social perspective taking (which entailed "walking in each other's shoes"). The perspective reversal training occurred 10 minutes daily during a free-play period over 9 consecutive school days for a total of 1.5 hours. Children who did not receive perspective reversal training also met daily during the free-play period for an identical amount of time to discuss friendship.

Additionally, children in the conflict-resolution-based-classroom-management conditions systematically used the six-step integrative negotiation procedure as the classroom management strategy for resolving real conflicts that occurred. The classroom teacher had children who experienced actual conflicts apply the integrative negotiation procedure to resolve their disputes. Eight of the 10 children in one of the classroom-management-use conditions applied the integrative negotiation procedure to resolve actual conflicts a total of 20 times; 6 of the 10 children in the other classroom-management-use condition applied the procedure a total of 18 times. Real use of the procedure, therefore, occurred in both conditions nearly equally over the same 9 consecutive school days as the role reversal training.

Dependent Measures

Eight dependent variables were measured and scored independently by two coders. All oral responses were tape recoded, transcribed verbatim, and scored using established rubrics. Interrater agreement on the *Interpersonal Negotiation Strategies Interview* scores was 93% on the pretest and 86% on the posttest. Interrater agreement on the scores of all other dependent measures was 98% or higher.

The first dependent variable was children's knowledge and understanding of the concept "conflict." Pre- and post-measures were obtained using the *What Conflict Means To Me* interview which asked, "Have you ever heard of the word conflict? What does it mean? How would you explain what conflict is to someone who does not know?" The scoring rubric for knowledge of conflict was 0 points for no knowledge, 1 point for accurately defining or providing an example of conflict, 2 points for accurately defining conflict and indicating that it should be resolved constructively.

The second dependent variable was children's learning of the integrative negotiation procedure. Pre- and post-measures were obtained using the *How I Manage Conflict* interview which asked, "When you have a conflict with someone, what do you do to solve it?" Responses were scored for the presence of the six steps of integrative negotiation (1 point each).

The third dependent variable was willingness and ability to apply the integrative negotiation procedure to conflict situations like those that occur in school. A post-measure was obtained using the *Computer Conflict Scenario* interview which asked what each child would do if blocked by a classmate from using the classroom computer. Responses were scored for the presence of the six steps of the negotiation procedure (1 point each).

The fourth dependent variable was children's willingness and ability to apply the integrative negotiation procedure in an actual conflict. A behavioral-observational measure was obtained 6 weeks after the study ended using the *Pair Conflict Task*, which consisted of each child individually making a May Day flower for a class bouquet. Children were paired within each condition, but did not have enough materials to make two complete flowers, and therefore faced a scarcity-of-resources conflict. Each pair was videotaped and scored for using integrative negotiation: 1 point total for either child recognizing the conflict and suggesting to solve it constructively, 1 point per child for expressing wants (2 points possible), 1 point per child for expressing feelings (2 points possible), 1 point per child for stating reasons (2 points possible), 1 point per child for communicating understanding (2 points possible), 1 point total for generating possible solutions, and 1 point total for mutually reaching an agreement. The maximum pair score was 11 points.

The fifth dependent variable was children's knowledge and understanding of feelings that people experience in conflict. Pre- and post-measures were obtained using the *Conflict Feelings Interview* which asked, "How do people feel when they are having a problem or conflict with someone? What are all of the feelings people might have when they are in a conflict?" The total number of different feelings listed was tallied (1 point each).

The sixth dependent variable was children's reasoning about the purpose, importance, and interrelatedness of each of the six steps in the integrative negotiation procedure. A post-measure was obtained using the *Integrative Negotiation Steps Interview* in which children could answer "*yes*" or "*no*" or "*not sure*" when asked, "Are any of the conflict steps more important than the others? Which ones? Why?" Responses were scored using the *Integrative Negotiation Reasoning Scale* which arranges reasons on a continuum from least to most complex: 0 points for "I don't know;" 1 point for "the first step stops the fighting and starts conflict resolution, the last step ends the conflict;" 2 points for "you cannot tell which steps are most important, all of the steps are the same;" 3 points for "all of the steps are linked, you need to use all of them to resolve a conflict;" 4 points for "wants, feelings, and reasons help the other person better understand your perspective;" 5 points for "wants, feelings, and reasons help you better understand the other person's perspective and communicate understanding to facilitate mutual problem solving."

The seventh dependent variable was children's developmental maturity and ability to reason about the process of interpersonal negotiation in conflict situations. Pre- and post-measures were obtained using the *Interpersonal Negotiation Strategies Interview* that

presented a scenario about two friends/classmates who are in conflict over using the only available red felt marker, followed by eight questions designed to assess social-cognitive development (see Schultz, Yeates, & Selman, 1989; Stevahn, Oberle, Johnson, & Johnson, 2001, April). Responses were coded and averaged as specified in The *Interpersonal Negotiation Strategies Interview Manual* (Schultz, Yeates, & Selman, 1989). Final scores range from 0 to 3 on Selman's (1980, 1981) developmental continuum: 0 reflects undifferentiated/egocentric/impulsive behavior (fighting or fleeing), 1 reflects differentiated/subjective/unilateral behavior (commanding or appeasing), 2 reflects self-reflective/reciprocal behavior (making fair trades and exchanges), 3 reflects a third-person/mutual orientation toward resolving conflict (problem-solving collaboratively in search of mutually satisfying outcomes).

Responses to the red marker conflict in the *Interpersonal Negotiation Strategies Interview* also were scored holistically using two behavior-categorization scales from previous studies (see Stevahn et al., 2000). The *Conflict Management Scale*, which arranges behaviors for dealing with conflict on a 12-point continuum from most destructive (physical aggression) to most constructive (full integrative negotiation), ranked the constructiveness of the children's proposed resolutions. The *Conflict Strategies Theory Scale*, derived from dual-concerns theory (see Johnson & Johnson, 2003; Johnson & Johnson, 1995), classified the proposed resolutions on the following 5-point continuum: forcing, withdrawing, smoothing, compromising, negotiating for mutual gain.

The eighth dependent variable was children's conceptual understanding of friendship. Pre- and post-measures were obtained using the *Friendship Interview*, which asked: "What does it mean to be a friend? What do friends do?" Responses were scored in two ways. The total number of different friendship concepts was tallied (1 point each) and responses that indicated "friends solve conflicts constructively" were tallied (1 point each).

Procedure

The study was conducted in Winter 2000 and comprised four phases: pretesting, training, posttesting, and follow-up testing. Five pre-measures were administered: *What Conflict Means To Me, How I Manage Conflict, the Conflict Feelings Interview*, the *Interpersonal Negotiation Strategies Interview*, and the *Friendship Interview.* The kindergartners in both the morning and afternoon classes were randomly assigned to receive the curriculum-integrated conflict resolution training either with or without role reversal training, and the afternoon and morning classes, respectively, were randomly assigned to either use or not use the integrative negotiation procedure to resolve actual conflicts for classroom management. The respective treatments were administered over 13 consecutive school days for identical amounts of time and used the same curriculum materials. Post-measures identical to the pre-measures were administered at the end of the study, along with two post-only measures: the *Computer Conflict Scenario Interview* and the *Integrative Negotiation Steps* interview. The Pair Conflict Task was administered 6 weeks later.

Analysis

A two-way analysis of variance was used to determine differences among conditions. When there was a significant interaction effect, *t* tests were used to interpret the results. When there were no significant differences, the four conditions were combined and paired t tests were used to compare pre-post mean scores.

Results

Regarding the first dependent variable, before training there were no significant differences between conditions on children's knowledge and understanding of the concept "conflict" (Table 1). After training, the definitions in the perspective-reversal-conflict-trained conditions were significantly more complete compared to the definitions in the conflict-trained-only conditions, $F(1, 34) = 4.859, p < .034$.

Regarding the second dependent variable, both before and after training, there were no significant differences among conditions on children's learning of the integrative negotiation procedure measured by the *How I Manage Conflict* interview (Table 2). None of the children knew any of the negotiation steps prior to training and almost all of the children knew all of the steps after training. When conditions were combined and pre-post means compared, there was a significant difference in learning of the integrative negotiation steps after training compared to before training, $t(37) = 44.499$, $p < .000$ (Table 3).

Regarding the third dependent variable, after training there was a significant interaction effect on children's willingness and ability to apply the integrative negotiation procedure to resolving a computer conflict in the perspective-reversal-conflict-trained-classroom-management-use condition (Table 2). Those children significantly used more of the integrative negotiation steps to resolve the conflict compared to the other three conditions, $F(1, 34) = 5.935$, $p < .02$. Furthermore, both before and after training, there were no significant differences among conditions for responses on the *Interpersonal Negotiation Strategies Interview* about a conflict over a red felt marker classified using the *Conflict Management Scale* (Table 4). Before training, all of the children primarily indicated that the conflict should be resolved by withdrawing in a positive way or by invoking social norms; after training, all of the children primarily indicated that integrative negotiation should be used to resolve the conflict. When conditions were combined and pre-post means compared, there was a significant difference in the children's application of the integrative negotiation procedure to resolve the conflict after training compared to before training, $t(37) = 9.181, p < .000$ (Table 3). Similarly, both before and after training, there were no significant

Table 1

Definition of Conflict: What Conflict Means To Me

Condition		Pretest		Posttest	
	n	*M*	*SD*	*M*	*SD*
Conflict-trained; no CM use	9	0.00	0.00	1.44	0.53
Perspective-reversal-conflict-trained; no CM use	9	0.11	0.33	2.00	0.00
Conflict-trained; CM use	10	0.00	0.00	1.40	0.52
Perspective-reversal-conflict-trained; CM use	10	0.20	0.63	1.50	0.53

Note. CM = classroom management. Pretest: no significant differences among conditions. Posttest: main effect in the perspective-reversal-conflict-trained conditions, $F(1, 34) = 4.859, p<.034$.

Table 2

Learning and Application of the Integrative Negotiation Procedure

Condition		Pretest		Posttest	
	n	*M*	*SD*	*M*	*SD*
How I Manage Conflict					
Conflict-trained; no CM use	9	0	0	5.78	0.67
Perspective-reversal-conflict-trained; no CM use	9	0	0	5.78	0.67
Conflict-trained; CM use	10	0	0	5.40	1.26
Perspective-reversal-conflict-trained; CM use	10	0	0	6.00	0.00
Computer Conflict Scenario					
Conflict-trained; no CM use	9			4.78	1.30
Perspective-reversal-conflict-trained; no CM use	9			3.22	3.07
Conflict-trained; CM use	10			4.10	2.85
Perspective-reversal-conflict-trained; CM use	10			6.00	0.00
Pair Conflict Task					
Conflict-trained; no CM use	3			10.00	1.00
Perspective-reversal-conflict-trained; no CM use	4			10.75	0.50
Conflict-trained; CM use	4			11.00	0.00
Perspective-reversal-conflict-trained; CM use	5			11.00	0.00

Note. CM = classroom management. *How I Manage Conflict* pretest and posttest: no significant differences among conditions (see Table 3 for combined pre-post comparison results). *Computer Conflict Scenario*: interaction effect in the perspective-reversal-conflict-trained-classroom-management-use condition, $F(1, 34) = 5.935, p<.02$. *Pair Conflict Task*: main effect in the classroom-management-use conditions, $F(1, 34) = 6.598, p<.025$.

Table 3

Pre-Post Mean Score Comparisons on Dependent Measures

Measure	Trained (posttest)	Untrained (pretest)	*t* test	Probability
How I manage conflict: recall of negotiation steps	5.74(0.79)	0	44.499(37)	<.000
INS interveiw: conflict management scale	10.58(2.33)	6.63(1.38)	9.181(37)	<.000
INS interview: conflict strategies theory scale	4.16(1.55)	1.87(1.26)	6.438(37)	<.000
INS interview: developmental levels	2.03(0.43)	1.42(0.32)	9.217(37)	<.000
Friendship interview: total friendship concepts	4.42(1.70)	2.95(1.74)	4.646(37)	<.000
Friendship interview: friends solve conflict responses	0.37(0.54)	0.003(0.16)	3.949(37)	<.000

Note. INS = Interpersonal Negotiation Strategies.

Table 4

Interpersonal Negotiation Strategies Interview: Red Felt Marker Conflict Scenario

Condition		Pretest		Posttest	
	n	*M*	*SD*	*M*	*SD*
Developmental Levels Scale					
Conflict-trained; no CM use	9	1.47	0.36	1.94	0.55
Perspective-reversal-conflict-trained; no CM use	9	1.52	0.31	2.07	0.42
Conflict-trained; CM use	10	1.32	0.36	2.01	0.50
Perspective-reversal-conflict-trained; CM use	10	1.38	0.27	2.08	0.28
Conflict Management Scale					
Conflict-trained; no CM use	9	6.44	1.81	10.00	3.12
Perspective-reversal-conflict-trained; no CM use	9	6.89	0.60	10.33	2.50
Conflict-trained; CM use	10	6.90	0.88	10.30	2.21
Perspective-reversal-conflict-trained; CM use	10	6.30	1.89	11.60	1.26
Conflict Strategies Theory Scale					
Conflict-trained; no CM use	9	2.11	1.36	3.67	2.00
Perspective-reversal-conflict-trained; no CM use	9	1.78	1.20	3.67	2.00
Conflict-trained; CM use	10	1.80	1.32	4.30	1.25
Perspective-reversal-conflict-trained; CM use	10	1.80	1.32	4.90	0.32

Note. CM = classroom management. For each scale on pretests and posttests there were no significant differences among conditions (for each scale see Table 3 for combined pre-post comparison results).

differences among conditions for the responses classified using the Conflict Strategies Theory Scale (Table 4). Before training, all of the children primarily indicated that the conflict should be resolved by withdrawing; after training, all of the children primarily indicated that the conflict should be resolved by smoothing, compromising, or using integrative negotiation. When conditions were combined and pre-post means compared, there was a significant increase in the children's use of constructive conflict resolution procedures after training compared to before training, $t(37) = 6.438, p < .000$ (Table 3).

Regarding the fourth dependent variable, 6 weeks after training there was a significant main effect in the classroom-management-use conditions on children's willingness and ability to apply the integrative negotiation procedure to resolve the actual conflict in the *Pair Conflict Task* (Table 2). The children in those conditions significantly used more of the integrative negotiation procedure compared to the children in the conditions lacking classroom management use of the procedure, $F(1, 34) = 6.598, p < .025$.

Regarding the fifth dependent variable, before training there were no significant differences among conditions in children's knowledge and understanding of conflict feelings (Table 5). After training, there was a significant main effect on children's conceptualizations of conflict feelings in the perspective-reversal-conflict-trained conditions. Those children listed significantly more feelings than children in the conflict-trained-only conditions, $F(1, 34) = 8.637, p < .006$.

Regarding the sixth dependent variable, after training there was a significant main effect on children's reasoning about the purpose/importance/interrelatedness of the integrative negotiation steps in the perspective-reversal-conflict-trained conditions (Table 6). Those children demonstrated significantly more sophisticated reasoning about negotiation steps compared to the children in the conflict-trained-only groups, $F(1, 34) = 9.998, p < .003$.

Regarding the seventh dependent variable, both before and after training, there were no significant differences among conditions on children's responses to the Interpersonal Negotiation Strategies Interview, which measured cognitive-social developmental growth (Table 4). At the start of the study children tended to be at the differentiated/subjective to self-reflective/reciprocal levels of development; at the end of the study children tended to be at the self-reflective/reciprocal level of development. When conditions were combined and pre-post means compared, there was a significant increase in the level of cognitive-social development after training compared to before training, $t(37) = 9.217, p < .000$ (Table 3).

Regarding the eighth dependent variable, both before and after training, there were no significant differences among conditions on children's conceptual understanding of friendship (Table 7). Before the study, all children knew approximately 3 friendship concepts, whereas after the study, all children knew 4 or 5 friendship concepts. When conditions were combined and pre-post means compared, there was a significant increase in children's total recall of friendship concepts, $t(37) = 4.646, p < .000$ (Table 3). Similarly, both before and after training, there were no significant differences among conditions on children's responses that indicated friends constructively resolve conflict (Table 7). Before the study, none of the children in any of the groups said that friends resolve conflicts, whereas after the study, several children in each group said that friends resolve conflicts. When conditions were combined and pre-post means compared, there was a significant increase in children's conceptual understanding about friends working together to resolve conflicts constructively, $t(37) = 3.949, p < .000$ (Table 3).

Table 5

Knowledge and Understanding of Conflict Feelings

Condition		Pretest		Posttest	
	n	*M*	*SD*	*M*	*SD*
Conflict-trained; no CM use	9	2.56	0.73	4.56	1.42
Perspective-reversal-conflict-trained; no CM use	9	2.33	1.12	6.33	2.00
Conflict-trained; CM use	10	2.40	1.26	4.00	1.70
Perspective-reversal-conflict-trained; CM use	10	2.20	1.03	5.30	1.25

Note. CM = classroom management. Pretest: no significant differences among conditions. Posttest: main effect in the perspective-reversal-conflict-trained conditions, $F(1, 34) = 8.637, p < .006$.

Table 6

Reasoning on the Purpose, Importance, and Interrelatedness of the Integrative Negotiation Steps

Condition		Posttest	
	n	*M*	*SD*
Conflict-trained; no CM use	9	2.78	1.48
Perspective-reversal-conflict-trained; no CM use	9	3.33	1.41
Conflict-trained; CM use	10	2.00	1.05
Perspective-reversal-conflict-trained; CM use	10	4.20	1.40

Note. CM = classroom management. Main effect in the perspective-reversal-conflict-trained conditions, $F(1, 34) = 9.998, p<003$.

Table 7

Academic Achievement: Conceptual Understanding of Friendship

Condition		Pretest		Posttest	
	n	*M*	*SD*	*M*	*SD*
Total friendship concepts					
Conflict-trained; no CM use	9	3.56	2.19	4.33	1.87
Perspective-reversal-conflict-trained; no CM use	9	3.00	2.12	5.00	2.45
Conflict-trained; CM use	10	2.60	1.26	4.00	1.33
Perspective-reversal-conflict-trained; CM use	10	2.70	1.42	4.40	1.07
Friends solve conflict responses					
Conflict-trained; no CM use	9	0.00	0.00	0.44	0.73
Perspective-reversal-conflict-trained; no CM use	9	0.00	0.00	0.22	0.44
Conflict-trained; CM use	10	0.01	0.32	0.30	0.48
Perspective-reversal-conflict-trained; CM use	10	0.00	0.00	0.50	0.53

Note. CM = classroom management. For total friendship concepts and friends solve conflict responses on pretests and posttests there were no significant differences among conditions (for each measure see Table 3 for combined pre-post comparison results).

Discussion

Teaching all learners how to resolve conflicts constructively has the potential to provide both immediate and long-term benefits to students, teachers, and schools. Children who learn conflict skills at an early age may get a "head start" in developing more positive peer relations, interpersonal competence, social adjustment, and foundational knowledge for future learning. Teachers who systematically incorporate conflict training into classroom management may more effectively deal with daily discipline issues, as well as provide students with ongoing opportunities to exercise self-responsibility and self-regulation. Schools in which everyone co-orients to a set of values and procedures for managing disputes constructively may become conflict-positive organizations, better able to establish safe and productive learning environments. Young children, however, often are not included in conflict resolution training programs. Devoting the time, energy, and resources to do so makes sense only if those children can learn what is taught.

The first issue in this study was the extent to which the school-based conflict resolution training for kindergartners

was effective. The findings clearly indicate that the training was effective on several levels, including enhancing children's ability to (a) distinguish conflict, (b) recall procedures for resolving it, (c) apply procedures in scenario and simulation situations, and (d) spontaneously use procedures to resolve actual conflicts with peers. Although knowing what conflict is, recognizing that it exists, and remembering steps for constructively resolving it preclude taking action to do so, the most convincing evidence of effectiveness is whether students skillfully apply the procedures when they face real conflicts. In this study, all of the children exhibited willingness and ability to use the integrative negotiation procedure to resolve the "real" conflict in the *Pair Conflict Task* that arose 6 weeks after the studied ended, varying only in the extent to they included all of the steps. The children appear to have internalized habit patterns for using the steps.

The second issue addressed in this study was the extent to which perspective reversal training and classroom management use of integrative negotiation mediated the effectiveness of the conflict training. The findings indicate that both of these factors increased the effectiveness of the conflict training. Specifically, perspective reversal training enhanced children's conceptualizations of conflict, resulting in more elaborate definitions of conflict that included the need for constructive resolution. Systematic classroom management use of the integrative negotiation procedure also positively affected children's willingness and ability to use integrative negotiation to resolve real conflict. When combined, these two factors interacted in a way that significantly and positively affected children's use of the procedure to resolve the computer conflict scenario. Although interaction effects did not occur on the "real use" measures, heightened emotions in actual conflict situations may inhibit one's ability to readily engage in social perspective taking without more practice. Even so, perspective reversal training did significantly and positively affect children's ability to identify feelings experienced in conflicts and children's reasoning about the purpose, importance, and interrelatedness of the steps in the integrative negotiation procedure. These findings are noteworthy because expressing feelings is fundamental to being able to resolve conflicts constructively, yet is difficult to do (Johnson & Johnson, 2003; Johnson & Johnson, 1995). Similarly, developing more complex conceptualizations about the purpose and functions of the integrative negotiation steps may be necessary for skillful use. Generating integrative solutions that maximize joint outcomes, for example, depends on clearly understanding the other person's wants, interests, feelings, and reasons as well as one's own. In this study, 10 of the 19 children who received the perspective reversal training noted the ultimate importance of social perspective taking and how that component gives purpose and meaning beyond self-interest to the act of expressing one's own wants, feelings, and reasons; none of the 19 children who did not receive perspective reversal training did so.

The third issue addressed in this study was the extent to which the conflict training affected children's cognitive-developmental capacity to engage in social perspective coordination in resolving conflicts. Some believe that children's conceptions about conflict and approaches to managing it are limited by chronological age and corresponding developmental growth that progresses gradually over time. Others believe that conflict training at early ages may provide children with a "developmental boost," helping them to learn more quickly what they ordinarily would not be able to learn until older. The findings in this study indicate that children's natural developmental responses to conflict can be enhanced through training. For example, before training, most children perceived the red felt marker conflict scenario as a scarcity-of-resources problem, illustrated by this child's response:

> The teacher should have gave them more markers because they were fighting and if someone let go the one would get it, and then the other wouldn't get a turn. They both wanted to color first. They should share it because then it's fair.

After training, that same child, like others, perceived the conflict as a relationship problem:

> I'd tell them to work it out—maybe use the conflict steps. They won't share the marker and that's a problem because if you don't share, your friendship breaks. They should take turns because then they can get along with friendship again. They'd both feel excited and happy and glad because they'd be friends again.

Resolving conflict by "making trades and exchanges to preserve fairness" is substantively different from "using conflict resolution to preserve friendship."

The fourth issue addressed in this study was the impact of the conflict resolution training on children's conceptual understanding of friendship, which was the academic achievement outcome. Despite the benefits of teaching children how to manage conflict constructively, academic achievement largely remains the central concern and focus of schooling. Demonstrating the feasibility of integrating conflict resolution training into academic programs and positive effects on student achievement, therefore, may be important to the implementation and institutionalization of such programs. The results of this study demonstrated both the feasibility and effectiveness of the integrated approach. The children expanded their conceptualizations of friendship both by exhibiting a broader general knowledge base and by associating conflict resolution with friendship. Before training, none of the children associated friendship with constructive conflict resolution. After training, however, 13 of the 38 children indicated that friends resolve conflicts constructively.

In summary, this study adds important new evidence

on the effectiveness of school-based conflict resolution programs. First, it provides new data on the effectiveness of teaching 5- and 6-year-olds to constructively manage conflicts in school settings. The behavioral-observational measure that assessed the willingness and ability of all of the kindergartners to use integrative negotiation in a real conflict especially demonstrated the children's competence in transferring what they learned. This corroborates findings in previous studies conducted on elementary, middle, and high school students in diverse school and community settings in both the United States and Canada (Johnson & Johnson, 2002), thereby strengthening the claim that the desired outcomes of conflict training generalize across a wide range of ages and contexts.

Second, this study provides new information on two factors that may mediate the effectiveness of conflict resolution training in schools. The first is perspective reversal training. Being able to effectively engage in social perspective taking is fundamental to integrative negotiations aimed at maximizing joint outcomes. The second is systematic classroom management use of the conflict procedure. Routinely using integrative negotiation to resolve real conflicts that occur in the classroom not only provides children with the type of practice that promotes transfer of training; it also helps establish social norms that support transfer of training. Although the children in this study who routinely used the negotiation procedure for classroom management also appeared to be better able to use it to resolve real conflicts that unexpectedly occurred, future studies need to provide further data on larger samples that receive treatment over longer periods of time.

Third, this study provides new data on the impact of conflict resolution training on children's cognitive-social development. This is the first school-based conflict training study to use a developmental measure to assess children's cognitive-social growth in managing conflicts. The results indicate that not only do young children have the capacity to learn and use more sophisticated interpersonal negotiation strategies, young children may gain a developmental advantage from doing so.

Fourth, this study validated the feasibility and effectiveness of both the curriculum-integrated approach to conflict training and the total-student-body approach to institutionalizing conflict programs in schools. Potentially all children in school can receive conflict training when it is incorporated into required subjects and curricular units. This enables everyone to actively participate in making schools safe, conflict-positive places. Such training especially makes sense when it simultaneously promotes academic learning. The findings in this study support the claim that conflict training and academic learning can be combined in ways that enhance both.

Finally, this study provides a link between theory and practice. Both the integrative negotiation procedure and the perspective reversal training were directly based on theorizing in the field of conflict resolution (Deutsch, 1973; Johnson, 1967, 1971, 1974; Johnson & Johnson, 1995; Pruitt, 1981). The results, therefore, build a bridge between the conflict theories primarily developed and tested in laboratory settings and the utility of those theories in "real-life" settings such as schools.

References

Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. New Haven, CT: Yale University Press.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers CollegePress.

Johnson, D. W. (1967). The use of role reversal in intergroup competition. *Journal of Personality and Social Psychology, 7*, 135-141.

Johnson, D. W. (1970). *The social psychology of education.* New York: Holt, Rinehart, & Winston.

Johnson, D. W. (1971). Role reversal: A summary and review of the research. *International Journal of Group Tensions, 1*, 318-334.

Johnson, D. W. (1974). Communication and the inducement of cooperative behavior in conflicts: A critical review. *Speech Monographs, 41*, 64-78.

Johnson, D. W., & Johnson, F. P. (2003). *Joining together: Group theory and group skills* (8th ed.). Boston: Allyn & Bacon.

Johnson, D. W., & Johnson, R. T. (1995). *Teaching students to be peacemakers* (3rd ed.). Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (1996). Conflict resolution and peer mediation programs in elementary and secondary schools: A review of the research. *Review of Educational Research, 66*(4), 459-506.

Johnson, D. W., & Johnson, R. T. (2002). Teaching students to be peacemakers: A meta-analysis. *Journal of Research in Education, 12*(1), 25-39.

Kohlberg, L. (1987). *Childhood psychology and childhood education: A cognitive developmental view.* New York: Longman.

Piaget, J. (1928). *Judgment and reasoning in the child.* London: Routledge & Kegan Paul.

Piaget, J. (1932). *The moral judgment of the child.* London: Routledge & Kegan Paul.

Pruitt, D. G. (1981). *Negotiation behavior.* New York: Academic Press.

Schultz, L. H., Yeates, K. O., & Selman, R. L. (1989). *The interpersonal negotiation strategies interview manual.* Graduate School of Education and Department of Psychiatry, Harvard University.

Selman, R. L. (1980). *The growth of interpersonal understanding: Developmental and clinic alanalyses.* New York: Academic Press.

Selman, R. L. (1981). The development of interpersonal competence: The role of understanding in conduct. *Developmental Review, 1*, 401-422.

Shantz, C. U. (1987). Conflicts between children. *Child Development, 58*, 283-305.

Stevahn, L., Oberle, K., Johnson, D. W., & Johnson, R. T. (2001, April). *Effects of role reversal training and use of integrative negotiation for classroom management on conflict resolution in kindergarten.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Stevahn, L., Johnson, D. W., Johnson, R. T., Oberle, K., & Wahl, L. (2000). Effects of conflict resolution training integrated into a kindergarten curriculum. *Child Development, 71*(3), 772-784.

Stevahn, L., Johnson, D. W., Johnson, R. T., & Real, D. (1996). The impact of a cooperative or individualistic context on the effectiveness of conflict resolution training. *American Educational Research Journal, 33*(3), 801-823.

Laurie Stevahn is Assistant Professor of Education and Director of Curriculum and Instruction at Seattle University. She specializes in social interdependence, conflict resolution, and organization development.

Katie Oberle is a National Board Certified Teacher candidate teaching kindergarten in Edina Public Schools, Minnesota. She specializes in creating cooperative classroom communities and developing early childhood literacy

The Impact of Academic Controversy on Subsequent Conflict Resolution and Relationships Among Students

James M. Mitchell
California State University, Hayward

David W. Johnson
and
Roger T. Johnson
University of Minnesota

The effects were studied of academic controversy versus individualistic learning on the constructive resolution of conflicts and the nature of relationships among students. For eight weeks participating students studied a leadership and citizenship curriculum using either academic controversy or individualistic learning. One-thousand-three-hundred-sixty-eight K-12 students in an inner-city school district participated in the study. A representative random sample of 176 students from grades 3, 4, 5, 9, and 11 was selected to be in the analysis. The results indicate that participating in academic controversies resulted in the use of more constructive conflict resolution strategies in future conflicts and improved the working relationship among students.

Many teachers avoid and suppress conflict among students during academic lessons. They tend to view conflict negatively, seeing it as interfering with learning and decreasing the quality of the learning environment. Yet there is evidence that conflict can have many positive outcomes, such as increasing motivation to learn and creative thinking (Deutsch, 1973; Johnson & F. Johnson, 2003). In training teachers in the use of effective instructional methods, the use of conflict to promote learning and enhance the learning environment is largely ignored. What is often taught is folklore and personal experiences rather than systematic programs based on theory and research. One of the few programs that is based on theory and research is academic controversy (Johnson & Johnson, 1979, 1995, 2000). Academic controversy exists when one individual's ideas, information, conclusions, theories, and opinions are incompatible with those of another and the two seek to reach an agreement (Johnson and Johnson, 1995). To engage in a controversy students must progress through the stages of (a) researching and preparing a position, (b) presenting and advocating their position, (c) engaging in a general discussion in which they refute the opposing position and rebut attacks on their own position, (d) reversing perspectives by presenting the best case for the opposing position, and (e) creating a synthesis on which all group members can agree. There is considerable evidence indicating that controversy, compared to concurrence-seeking discussions, debate, and individualistic study, tends to create higher achievement, greater retention, greater creativity, and more critical thinking and use of higher-level reasoning (Johnson & Johnson, 1979, 1995, 2000).

There are at least three shortcomings of the current research on controversy. The first is that there is no evidence as to whether the procedures and skills learned by engaging in academic controversies results in students being able to manage "real world" conflicts in a more constructive way. There is evidence that the more skillful participants are in engaging in the controversy procedure, the more positive the outcomes (Johnson & Johnson, 1989, 1995). Practice seems to make perfect. Thus, engaging in academic controversies tends to increase the competencies of participants to engage in future controversies. It is unknown, however, whether these skills will transfer to actual conflicts among students. The first purpose of this study, therefore, is to investigate this issue.

The second shortcoming is that there is very little evidence concerning the impact of engaging in academic controversies on the working relationships of participants. It is often assumed that conflict results in rejection, divisiveness, and hostility among participants (Collins, 1970). On the other hand, conflicts have been hypothesized potentially to strengthen relationships and increase the ability of individuals to work effectively with each other (Deutsch, 1962; Johnson, 1970; Johnson & Johnson, 1989, 1995). Working relationships among students include seeking out each other's conclusions and helping each other

learn. There is, however, very little evidence as to whether conflict decreases or increases the working relationship of disputants. In this study, therefore, students' commitment to each other's learning was investigated.

The third shortcoming of the current research on academic controversy is that its generalizability is limited by the setting, the type of participants, and the number of teachers involved. The previous research has primarily consisted of small experimental studies conducted in research laboratories and suburban schools. The number of experimenters or teachers involved has typically been two or at the most four. There is a need to conduct field studies in which numerous teachers use academic controversy to determine if the procedure is effective in a large-scale implementation. In addition, almost all of the previous research has been conducted in suburban school districts. There is a need to conduct studies in urban schools. Finally, the previous research has been conducted primarily with participants from European-American backgrounds. There is almost no controversy research utilizing minority participants. In this study, therefore, an inner-city, primarily minority student population was studied in a large scale implementation involving over fifty teachers.

A fourth issue examined in this study was the impact of teaching a series of academic controversies focused on citizenship on the decision of undergraduate students to enroll in a teacher-training program. Undergraduate students were able to take a year-long course on teaching to determine whether or not they wished to become teachers. The course included a teaching experience in an inner-city school. The intent of the course was to recruit capable undergraduate students to become teachers by giving them a direct public school teaching experience. The impact of such courses on the decision to become a teacher is largely unevaluated. Thus, the impact of the experience of teaching in an inner city school on whether the undergraduate students applied for admission to the teacher-training program was investigated.

Methods

Participants

A representative random sample of 176 students was selected from a population of 1,368 K-12 students from a large inner-city school district (60 percent were Black Americans, 15 percent were Asian, 15 percent were Latino, and 10 percent were European-American; there were approximately an equal number of males and females in the study). Trained teachers taught 932 students using the controversy procedure and 436 using the in the control condition. Of the 176 participants, 117 were in the controversy condition and 59 were in the control condition. The sample was drawn from grades 3, 4, and 5 and grades 9 and 11.

Fifty seven college undergraduate students served as teachers in the controversy condition. The undergraduates were 92 percent European-American, 4 percent African-American, and 4 percent Asian-American. Sixty-three percent were female and 37 percent were male. They were all interested in becoming teachers but had not yet committed themselves to the teacher-training program at the university. This class was aimed at recruiting undergraduate students to be teachers by giving them a direct teaching experience in an actual public school. Since the schools were in the inner-city, students could receive credit for participating in a community service project by taking the class.

The 37 individualistic learning control classes were taught by their regular teachers. The teachers were 60 percent European-American, 20 percent African-American, 10 percent Asian-American, and 10 percent Latino-American. Ninety-percent were female and 10 percent were male. The teachers were representative of all the teachers in the school district.

Independent Variable

The independent variable was the use of academic controversy versus the use of individualistic learning. *Academic controversy* was operationally defined as the implementation of the five-step controversy procedure (Johnson & Johnson, 1995). More specifically, students were randomly assigned to groups of four, which were then divided into two pairs. Each pair was randomly assigned either a pro or a con position on an issue being studied. For example, in one controversy, students were asked to determine which of two characters in a reading assignment was the better person to mediate a conflict. Although wholly different from one another, each character had strengths at conflict management. One pair was given the assignment for making the best case possible that Character A should be the mediator and the other pair was assigned Character B. The pairs (a) researched and prepared their position, (b) made a persuasive presentation to the opposing pair, (c) engaged in an open discussion in which they continued to advocate their position while attempting to refute the opposing position, (d) reversed perspectives by presenting the best case for the opposing position, and (e) jointly created a synthesis on which all four group members could agree and which represented the students' best reasoned judgment on the issue (Johnson and Johnson, 1995).

Individualistic learning was operationally defined as the use of the traditional lecture, whole class discussion, and individual worksheet format. Students worked alone and were assigned grades on a criteria-referenced basis.

Dependent Measures

The *Conflict Scenario Written Measure* was given to all participating students the day before the controversy program began and the day after the program ended. The measure consisted of a brief scenario that ended in an unresolved conflict and instructions for the student to write down what he or she would do if actually in the situation. Students' responses were scored on a five-point scale consisting of: (1) physical aggression or verbal threats, (2)

withdrawal, (3) attempting to force the other to give in by commanding him or her to do so or asking the teacher to force the other to concede, (4) compromising, or (5) creative problem-solving to reach an agreement that maximizes the benefits to both parties (Johnson and Johnson, 2000).

The *Peer Facilitation of Learning Scale* consisted of six Likert-type items, each of which was scored on a continuum from one to five (Johnson and Johnson, 1997). The items measured students' perceptions of their commitment to help classmates learn and seek out classmates' ideas and reasoning to enhance their own learning.

Commitment to teaching was measured by the number of students applying for admission to the teacher training program.

Procedure

Fifty-seven undergraduate college students were trained by a university professor to teach a curriculum unit using the instructional procedure of academic controversy. Regular classroom teachers were instructed to teach the same curriculum using individualistic learning. The college students attended an instructional methods course developed to recruit undergraduate students into teacher education by providing them with an actual classroom teaching experience in a public school. The undergraduate students were taught how to structure and conduct controversies, monitor the students as the students engaged in the controversy procedure, intervene when necessary to improve individual and team work, and evaluate the resulting academic achievement. The class lasted for three quarters (one academic year). Each quarter lasted for 10 weeks during which the class met twice a week for a total of 30 hours. The first quarter focused on the nature of the Citizenship Development Curriculum (aimed at teaching students general leadership and citizenship skills) and the nature of the academic controversy procedure. The second quarter consisted of the actual teaching experience—one class session was held at the university and one class session (for eight weeks) consisted of the undergraduates teaching the Citizenship Development Curriculum (utilizing the academic controversy procedure) to students in inner-city schools. The third quarter focused on analyzing the undergraduates' teaching experiences and discussing the nature of a career in teaching.

The eight lessons taught during the second quarter of the course consisted of the following activities. Two university undergraduates were assigned to each inner-city class. The eight class sessions consisted of four controversies (e.g., choosing the best mediator to resolve a conflict and choosing between a friend or a stranger to hire for a job based on their different qualifications), each lasting for two class sessions. Each class session lasted for fifty minutes. On the first class session the citizenship development issue was presented, sides were assigned, students prepared the best case possible for their position, and presented it to the opposing pair. The second class session consisted of students engaging in an open discussion of the issue in which they critically analyzed and refuted the opposing position, reversed perspectives and presented the best case for the opposing position, came to agreement as to their best reasoned judgment about the issue, and listened to the undergraduate teachers summarize and bring closure to the issue. The lessons were taught from kindergarten through the twelfth grades. Two forms of the unit were taught—one for kindergarten through second grades and one for third through twelfth grades.

The academic controversy procedure complemented the Citizenship Development Curriculum. Thomas Jefferson, James Madison, and other founders of the American Republic believed that political decisions should result from a free and open discussion characterized by conflict among ideas and opinions. They expected that the clash of opposing positions within political discourse would increase citizens' understanding of the issue and the quality of their collective decision making. Jefferson noted, "Differences of opinion lead to inquiry, and inquiry to truth." Within political discourse, each alternative course of action was expected to (a) be strongly advocated, (b) receive a complete and fair hearing, and (c) be critically analyzed to reveal its strengths and weaknesses. By integrating the academic controversy procedure into a unit on citizenship development, citizenship skills were taught both by the content and by the instructional procedures.

Analyses

A pre-post, control group design was employed in the study. A two-way, repeated measures ANOVA was used to determine results for the dependent measures of conflict resolution and peer facilitation of learning.

Results

From Table 1 it may be seen that compared with the untrained students, the trained students in Grades 3, 4, and 5 used more constructive procedures to resolve a conflict (such as compromising and creative problem-solving), $F(3,210) = 246.06, p < .0001$, as did the trained students in Grades 9 and 11, $F(3,134) = 147.97, p < .0001$.

In Grades 3, 4, and 5, the trained students (compared to untrained students) perceived more commitment to help classmates learn and greater motivation to seek out classmates' ideas and reasoning to enhance their own learning, $F(3,210) = 198.29, p < .0001$. The findings for Grades 9 and 11 were similar, $F(3,134) = 107.64, p < .0001$.

In September, at the beginning of the course, 40 of the 57 students indicated a desire eventually to enter the Teacher Education program. Of those 40 students, five decided that teaching would not be their chosen profession upon completion of the program in June. These students all expressed positive remarks about how glad they were to have had the opportunity to investigate the teaching profession before they would be required to invest the time

Table 1

Mean Responses To Conflict Scenario

Measure	*N*	Pre Mean	Pre *SD*	Post Mean	Post *SD*
Grades 3, 4, 5					
Conflict Trained	73	1.66	1.12	4.55	0.58
Untrained	34	1.27	0.62	1.09	0.52
Grades 9, 11					
Conflict Trained	44	1.64	1.08	4.46	0.63
Untrained	25	1.24	0.66	1.12	0.60

Note. Grades 3, 4, 5: $F(3,210) = 246.06, p < .0001$; Grades 9, 11: $F(3,134) = 147.97, p < .0001$

Table 2

Mean Responses To Peer Facilitation of Learning Scale

Measure	*N*	Pre Mean	Pre *SD*	Post Mean	Post *SD*
Grades 3, 4, 5					
Conflict Trained	73	1.64	1.12	4.49	0.56
Untrained	34	1.35	0.73	1.27	0.86
Grades 9, 11					
Conflict Trained	44	1.61	1.08	4.36	0.57
Untrained	25	1.36	0.81	1.34	1.00

Note. Grades 3, 4, 5: $F(3,210) = 198.29, p < .0001$; Grades 9, 11: $F(3,134) = 107.64, p < .0001$

Table 3

Undergraduates Entering Teacher Education Program

	N	Did Not Enter Teacher Education	Entered Teacher Education
Not Planning To Be Teachers	17	7	10
Planning To Be Teachers	40	5	35

and financial resources required by a full degree track. Likewise, the 35 remaining participants who were predisposed to the teaching profession have all expressed praise for this program as they enter their teacher education tracks. Of the 17 students who were primarily interested in the community service piece of the project, six left the program in January, one left in April, and ten decided to pursue teaching as a career. The six who left in January cited personal reasons for leaving the program, while the student who left in April cited geographic re-location as the reason. The ten students who decided to enter the teaching profession have all expressed positive statements about having had the experience.

Finally, of the original 57 teachers in the experimental condition, 42 stated that they learned a great deal about conflict resolution as a result of using the controversy procedure to teach public school students.

Discussion

While there are powerful instructional effects of using intellectual conflict in instructional situations, many teachers are reluctant to do so, perhaps because of a fear of conflict, a lack of knowledge of the procedures, or norms discouraging the use of conflict. To many teachers the controversy procedure may seem too risky or complex to utilize. The results of this study indicate that the controversy procedure is simple and safe enough that with a minimum of training, inexperienced undergraduates can use it successfully. The results are all the stronger as experienced teachers taught the individualistic condition and undergraduate students with no previous teaching experience taught the controversy condition.

There is considerable previous evidence that the use of academic controversy will enhance learning and higher-level reasoning (Johnson & Johnson, 1979, 1989, 1995, 2000). There are, however, significant gaps in the research that need to be explored. The purpose of this article was to explore three such gaps.

First, when teachers use academic controversy to teach academic content they do so with the hope that students will learn how to manage subsequent conflicts more constructively as well as learn in the current situation. Yet the impact of participating in academic controversies on subsequent behavior in conflict situations has never been investigated. The results of this study indicated that when given a description of a conflict and asked to describe what they would do to manage it, students who participated in academic controversies described a more effective strategy for managing the conflict (such as compromising and creative problem-solving) than did the students in the individualistic condition. This is an important added benefit to engaging students in academic controversies.

There is some evidence that friends rely more on negotiation and less on coercion than do acquaintances (Lauren, Hartup, & Koplas, 1996). Yet a common perception of conflict is that it is a divisive influence creating distrust, avoidance, and an unwillingness to work with the opponent in the future (Collins, 1970). Many teachers try to suppress conflict as they fear it will result in students trying to obstruct each other's learning. The results of this study indicate that engaging in academic controversies actually resulted in students having stronger commitment to helping classmates learn as well as greater motivation to seek out classmates' ideas and reasoning to enhance their own learning. This finding extends the known outcomes of academic controversy and strongly supports the position that conflict has positive outcomes when it is managed constructively.

The previous research involved a limited number of teachers or experimenters, primarily involved suburban students from a European-American background, and involved only one or two controversies. This study involved a large number of teachers in inner-city schools with mostly minority students and involved four different controversies. The results, therefore, are important as they increase the generalizability of the research on academic controversy.

Finally, a recent innovation in many colleges of education is to recruit into teaching undergraduate students who have not made a career choice via a class in which they have actual teaching experiences. There is, however, little evidence that such classes do result in more students enrolling in teacher education programs. The results of this study provide some evidence that the class studied was effective in influencing undergraduate students to choose a career in teaching.

When academic controversy is used, it is done so in the expectation that not only will there be immediate benefits for students such as greater learning but also that there will be long-term improvements in the ways in which students manage conflicts. The results of this study provide the first evidence that such as expectation is realistic and justified. In addition, while considerable time is spent in schools discussing and trying to repair relationships damaged by destructively managed conflicts, there is almost no evidence as to the importance of constructively managed conflicts in improving relationships and increasing students' ability to work together. The results of this study provide such evidence. The generalizability of the results of the research on academic controversy was limited by the small number and ethnicity of participants and the setting of the studies. The results of this study extend the generalizability of the knowledge about controversy. Finally, the study provides evidence that courses aimed at recruiting students to enter the teaching profession by giving them teaching experience are successful. Practitioners interested in effective teaching are, therefore, well advised to use academic controversies. The results are limited by the fact that novice teachers taught the experimental classes and experienced teachers taught the control classes, the inherent variability involved in a large number of teachers implementing the experimental conditions, and the paper and pencil nature of the dependent variables. Further research is needed to further extend the generalizability of the research on controversy, corroborate

the results of this study, and continue to extend the nature of the dependent variables studied.

References

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. Jones (Ed.), *Nebraska symposium on motivation* (pp. 275-319). Lincoln, NE: University of Nebraska Press.

Deutsch, M. (1973). *The resolution of conflict.* New Haven, CT: Yale University Press.

Collins, B. (1970). *Social Psychology.* Reading, MA: Addison-Wesley.

Johnson, D. W., & Johnson, R. (1979). Conflict in the classroom: Academic controversy and learning. *Review of Educational Research, 49,* 51-61.

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. (1995). *Creative Academic controversy: Intellectual challenge in the classroom.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. (1996a). Conflict resolution and peer mediation programs in elementary and secondary schools: A review of the research. *Review of Educational Research, 66*(4), 459-506.

Johnson, D. W., & Johnson, R. (2000). Academic controversy: the value of intellectual opposition. In M. Deutsch & P. Coleman (Eds.), *The handbook of conflict resolution* (pp. 65-85). San Francisco: Jossey-Bass.

Kaufman, P., Chen, X., Choy, S., Chandler, K., Chapman, C., Rand, M., & Ringel, C. (1999). *Indicators of school crime and safety, 1999.* Washington, D. C.: U.S. Departments of Education and Justice. NCES 999-251/NCJ-172215.

Laursen, B., Hartup, W., & Koplas, A. (1996). Towards understand peer conflict. *Merrill-Palmer Quarterly, 42,* 76-102.

James Mitchell is Assistant Professor of Teacher Education at California State, Hayward. His research interests focus on blending AIDS education and service learning within the frameworks of cooperative learning and conflict resolution

David W. Johnson is Professor of Educational Psychology at the University of Minnesota. He specializes in cooperation and competition, conflict resolution, and organizational change.

Roger T. Johnson is a Professor of Education at the University of Minnesota. He specializes in science education.

Author Note

This program was funded by the generous participation of the General Mills Foundation and the President of the General Mills Foundation, Dr. Reatha Clark King.

Controversy and Peace Education

David W. Johnson
and
Roger T. Johnson
University of Minnesota

Peace is freedom from war or strife. Peace education includes (a) facing the difficult issues (often involving ethnic, cultural, or religious differences) that must be discussed in order for peace to be established and maintained, (b) establishing a procedure (such as constructive controversy) that all parties agree to use to discuss the issues involved in the conflict, (c) training students how to use the procedure skillfully, and (d) incorporating the use of the procedure into students' personal identity so that it will be habitually used. In well structured controversies, participants make an initial judgment, present their conclusions to other group members, are challenged with opposing views, become uncertain about the correctness of their views, actively search for new information and understanding, incorporate others' perspectives and reasoning into their thinking, and reach a new set of conclusions. This process results in significant increases in the quality of decision making and problem solving (including higher-levels of cognitive and moral reasoning, perspective taking, creativity, and attitude change about the issue), motivation to learn more about the issue, positive attitudes toward the controversy and decision making processes, the quality of relationships, and self-esteem. The educational use of academic controversy may be utilized in any grade level and in any subject matter. The academic controversy procedure is a Level 3 intervention aimed at creating a tipping point in which enough individuals become skillful in the constructive controversy procedure that they will use it to establish and maintain peace. The possibility of academic controversies being structured throughout educational systems is strengthened by the foundation of theory and research on which the controversy procedure is based.

Nature Of Peace

In order to understand the nature of peace, it is necessary to understand the interrelationships among war, peace, cooperation, and conflict. *War* is a state of open and declared armed combat between states or nations, *peace* is freedom from war or strife, or a state of mutual concord between governments, cooperation is working together to achieve mutual goals, and conflict is the occurrence of incompatible activities (Johnson & Johnson, 1989, 1995a, 1995b; Johnson, Johnson, & Holubec, 2002). War and peace are two ends of a single continuum. If there is war, there is no peace and vice versa. Peace exists when there is cooperation among nations and war ends when cooperation is reestablished. Peace, however, is not an absence of conflict. Peace is a state in which conflicts occur frequently and are resolved constructively (war, in contrast, is a state in which conflicts are managed through the use of large scale violence). Conflicts should occur frequently because when they are managed constructively they have many positive outcomes, such as increasing motivation and energy to solve problems, increasing achievement and productivity, clarifying one's identity and values, and increasing understanding of other perspectives (Johnson, 1970; Johnson & Johnson, 1995a, 1995b). Especially in order to create and maintain peace, conflicting ideas, conclusions, and ideologies must be discussed so that shared understandings and mutual accommodations may take place.

Peace Education

One hope to establish and maintain peace is peace education. *Peace education* may be defined as teaching what peace is, how it may be established, how it may be maintained, and the factors influencing its demise. Peace

education includes (a) facing the difficult issues that must be discussed in order for peace to be established and maintained, (b) establishing a procedure that all parties agree to use to discuss these difficult issues, (c) training students how to use the procedure skillfully, and (d) incorporating the use of the procedure into students' personal identity and value system so that the procedure will be habitually used. Thus, peace education is directly aimed at cognitive knowledge, procedural knowledge, and affective states. The ultimate goal of peace education is to give students the knowledge, procedural competencies, identity, and values required to maintain peace within themselves (intrapersonal peace), among individuals (interpersonal peace), among groups (intergroup peace), and among countries, societies, and cultures (international peace).

Difficult Issues

In establishing and maintaining peace there are usually difficult issues and problems that have to be resolved. Problems that may result in war or violence are presented that involve ethnic differences, cultural differences, religious differences, economic interests, and political issues such as whether all members of a society are of equal worth or whether democracy, monarchy, or dictatorship is the best political system. In many of these issues personal identity is involved and certain courses of action either enhance that identity or contradict it. While complex issues such as these involve conflicts of interests (whose resolution requires negotiation, mediation, and arbitration), they also involve problem solving and decision making (which requires the use of the constructive controversy procedure).

The difficulty in discussing difficult issues is illustrated in a study conducted by Trosset (1998). She conducted a study at Grinnell College (a small and somewhat isolated college in Iowa) that has in the past two decades strived to increase the diversity of its student body. Over several semesters she presented approximately 200 students with a list of sensitive diversity-related issues (such as "whether race is an important difference between people") and asked (a) whether it was possible to have a balanced discussion of the issue (involving more than one perspective, with each perspective receiving about equal support and with people being civil to each other) and (b) why they did or did not want to discuss the issue. The majority of the students believed that balanced discussion of these issues was impossible because a single viewpoint would dominate and if anyone spoke against that perspective there would be reprisals. The students who wanted to discuss a particular topic typically held strong views on the subject and wished to convince others (i.e., their motivation was advocacy, not learning). The students who did not want to discuss an issue (a) tended not to have strong views ("*I don't know much about it, so I don't want to discuss it*") and (b) found the issue difficult to discuss ("*I never know how to approach this subject*"). Only 5 out of the 200 students stated they would like to discuss a difficult topic in order to learn more about it. Seventy-five percent of the students said they would discuss diversity issues with people of the same views or background as themselves, but only 40 percent said they would discuss the same issues with people whose views were unknown to them. A majority said they would be unlikely to listen to someone with whom they disagreed ("*I wouldn't want to participate in a conversation when other people have disagreeable views, but I would talk with people who have similar opinions.*"). One reason for avoiding discussions in which different opinions and conclusions were discussed was that the students believed the goal of such discussions was to reach consensus, rather than just learn from each other. From this study it may be concluded that even under ideal conditions, it is difficult for individuals to discuss major issues that affect conflict and peace among individuals.

In order to resolve these difficult issues, individuals need an effective procedure for doing so and build a personal identity based on "how" they deal with difficult issues rather than "what" their position is on these issues.

Identity

Discussing the difficult issues involved in establishing and maintaining peace may enhance or threaten personal identity. Social identity is derived from group membership, and group membership may be based on (a) beliefs and ideology or (b) how difficult issues are discussed. "What" a person thinks often determines his or her identity. A person cannot be a Christian or Hindu without holding a certain set of beliefs. If a person wants to be a member of certain groups, he or she must adopt the correct set of beliefs. If a person's beliefs are tied to his or her identity (that is, "what" he or she believes), the person will be reluctant to discuss the issue and resistant to changing his or her position. Defensively, the person may refuse to listen to or comprehend opposing views, as they may be perceived as personal attacks that threaten his or her identity. Changing beliefs may mean changing groups and identities. Under these conditions, open-minded discussion may be rare. Most discussions, therefore, are held with like-minded individuals. Expertise is assumed to depend on personal experiences as a member of the relevant group, which cannot be questioned or challenged. Thus, students of color are more likely than whites to claim to be knowledgeable about ethnicity and women are more likely than men to claim knowledge about gender. The bias towards personal experience as expertise (a) forces members of less powerful groups into the role of peer instructors and (b) supports the impression that members of more powerful groups have nothing legitimate to say.

Social identity may also be based on being an "educated person" who explores ideas from a variety of perspectives, learns about things outside his or her own experiences, evaluates the quality of evidence and arguments, and has the capacity to be persuaded of new perspectives when presented with high quality evidence and argument. The focus of educated individuals is on evidence and argument (as opposed to what they currently believe), since their

conclusions will always change in the face of new evidence and better arguments. In educated persons, identity tends to come from the process of how one thinks rather than what one thinks. Discussions will be sought out as a means for learning more about the issue, especially when someone disagrees.

Controversy And Peace Education

> Since the general or prevailing opinion on any subject is rarely or never the whole truth, it is only by the collision of adverse opinion that the remainder of the truth has any chance of being supplied.
>
> John Stuart Mill

In order to face and discuss constructively the difficult issues that are involved in establishing and maintaining peace (as opposed to war), it is necessary for individuals to have an agreed-upon procedure to use to structure the discussion and to have some expertise in using it. In addition, the use of the procedure may represent an identity based on being an educated person that supercedes identities based on group *membership* and certain beliefs. The procedure for decision making and problem solving that involves difficult issues is the constructive controversy procedure.

Constructive controversy exists when one person's ideas, information, conclusions, theories, and opinions are incompatible with those of another, and the two seek to reach an agreement (Johnson & Johnson, 1979, 1989, 1995b, 2000). Constructive controversy involves what Aristotle called *deliberate discourse* (i.e., the discussion of the advantages and disadvantages of proposed actions) aimed at synthesizing novel solutions (i.e., *creative problem solving*). Constructive controversy is most commonly contrasted with concurrence seeking, debate, and individualistic decisions. *Concurrence seeking* occurs when members of a group inhibit discussion to avoid any disagreement or argument, emphasize agreement, and avoid realistic appraisal of alternative ideas and courses of action. Concurrence seeking is close to Janis' (1982) concept of *groupthink* (i.e., members of a decision-making group set aside their doubts and misgivings about whatever policy is favored by the emerging consensus so as to be able to concur with the other members and thereby preserve the harmonious atmosphere of the group). *Debate* exists when two or more individuals argue positions that are incompatible with one another and a judge declares a winner on the basis of who presented their position the best. *Individualistic decisions* exist when individuals consider the issue alone while perceiving their goals to be unrelated and independent from the goals of others.

Theory Of Constructive Controversy

> There is no more certain sign of a narrow mind, of stupidity, and of arrogance, than to stand aloof from those who think differently from us.
>
> Walter Savage Landor

In an English class participants are considering the issue of civil disobedience. They learn that in the civil rights movement, individuals broke the law to gain equal rights for minorities. In numerous literary works, such as Huckleberry Finn, individuals wrestle with the issue of breaking the law to redress a social injustice. Huck wrestles with the issue of breaking the law in order to help Jim, the run-away slave. In the 1970s through 2003, prominent public figures from Wall Street to the White House have felt justified in breaking laws for personal or political gain. In order to study the role of civil disobedience in a democracy, participants are placed in a cooperative learning group of four members. The group is given the assignment of reaching their best reasoned judgment about the issue and then divided into two pairs. One pair is given the assignment of making the best case possible for the constructiveness of civil disobedience in a democracy. The other pair is given the assignment of making the best case possible for the destructiveness of civil disobedience in a democracy. In the resulting conflict, participants draw from such sources as the Declaration of Independence by Thomas Jefferson, Civil Disobedience by Henry David Thoreau, Speech at Cooper Union, New York by Abraham Lincoln, and Letter from Birmingham Jail by Martin Luther King, Jr. to challenge each other's reasoning and analyses concerning when civil disobedience is, or is not, constructive.

The teacher then follows the structured academic constructive controversy procedure over several class periods. The student pairs research the issue, prepare a persuasive case for their position, present their position in a compelling and interesting way, refute the opposing position while rebutting criticisms of their position, take the opposing perspectives, and derive a synthesis or integration of the positions. In conducting the constructive controversy, the teacher is operationalizing the theoretical process by which constructive controversy is implemented.

Over the past 35 years, we have (a) developed a theory of constructive controversy, (b) validated it through a program of research, (c) operationalized the validated theory into a practical procedure (there are two formats, one for decision-making situations and one for academic learning), (d) trained teachers, professors, administrators, managers, and executives in how to implement the constructive controversy procedure, and (e) developed a series of curriculum units, academic lessons, and training exercises structured for controversies. Our theorizing began with concepts taken from developmental, cognitive, social, and organizational psychology. The process through which constructive controversy creates positive outcomes involves the following theoretical assumptions (Johnson & Johnson, 1979, 1989, 1995b, 2000) (see Figure 1).

Table 1

Constructive Controversy, Debate, Concurrence-Seeking, and Individualistic Processes

Constructive Controversy	Debate	Concurrence-Seeking	Individualistic
Categorizing and Organizing Information to Derive Conclusions	Categorizing and Organizing Information to Derive Conclusions	Categorizing and Organizing Information to Derive Conclusions	Categorizing and Organizing Information to Derive Conclusions
Presenting, Advocating Elaborating Position and Rationale	Presenting, Advocating, Elaborating Position and Rationale	Presenting, Advocating, Elaborating Position and Rationale	No Oral Statement of Positions
Being Challenged by Opposing Views Results in Conceptual Conflict and Uncertainty about Correctness of Own Views	Being Challenged by Opposing Views Results in Conceptual Conflict and Uncertainty about Correctness of Own Views	Being Challenged by Opposing Views Results in Conceptual Conflict and Uncertainty about Correctness of Own Views	Presence of only One View Results in High Certainly About the Correctness of Own Views
Epistemic Curiosity Motivates Active Search for New Information and Perspectives	Closed Minded Rejection of Opposing Information and Perspectives	Apprehension About Differences and Closed-Minded Adherence to Own Point of View	Continued High Certainty About the Correctness of Own
Reconceptualization, Synthesis, Integration	Closed-Minded Adherence to Own Point of View	Quick Compromise to Dominant View	Adherence to Own Point of View
High Achievement, Positive Relationships, Psychological Health	Moderate Achievement, Relationships, Psychological Health	Low Achievement, Relationships, Psychological Health	Low Achievement, Relationships, Psychological Health

Note: Reprinted with permission from Johnson, D. W., & Johnson, R. (1995). *Creative controversy: Intellectual conflict in the classroom.* Edina, MN: Interaction Book Company.

Conditions Determining the Constructiveness of Controversy

1. When individuals are presented with a problem or decision, they have an initial conclusion based on categorizing and organizing incomplete information, their limited experiences, and their specific perspective. They have a high degree of confidence in their conclusions (they freeze the epistemic process).

2. When individuals present their conclusion and its rationale to others, they engage in cognitive rehearsal, deepen their understanding of their position, and use higher-level reasoning strategies. The more they attempt to persuade others to agree with them, the more committed they may become to their position.

3. When individuals are confronted with different conclusions based on other people's information, experiences, and perspectives, they become uncertain as to the correctness of their views and a state of conceptual conflict or disequilibrium is aroused. They unfreeze their epistemic process.

4. Uncertainty, conceptual conflict, or disequilibrium motivates *epistemic curiosity*, an active search for (a) more information and new experiences (increased specific content) and (b) a more adequate cognitive perspective and reasoning process (increased validity) in hopes of resolving the uncertainty.

5. By adapting their cognitive perspective and reasoning through understanding and accommodating the perspective and reasoning of others, individuals derive a new, reconceptualized, and reorganized conclusion. Novel solutions and decisions that, on balance, are qualitatively better are detected. The positive feelings and commitment individuals feel in creating a solution to the problem together is extended to each other and interpersonal attraction increases. Their competencies in managing conflicts constructively tend to improve. The process may begin again at this point or it may be terminated by freezing the current conclusion and resolving any dissonance by increasing the confidence in the validity of the conclusion.

The process that results, therefore, involves having an initial conclusion as to what course of action should be adopted to solve the problem, presenting a persuasive case for that conclusion while listening to persuasive presentations of opposing positions, feeling uncertain about the correctness of one's position, engaging in a search for better information and reconceptualizing one's views on the decision, and then coming to a new conclusion about what course of action should be adopted. Each time a person goes through this process his or her conclusions may be closer and closer approximations of the "truth."

Conditons Determining the Constructiveness of Controversy

Depending on the conditions under which controversy occurs and the way in which it is managed, controversy may result in positive or negative consequences. These conditions include the context within which the constructive controversy takes place, the level of group members' social skills, and group members' ability to engage

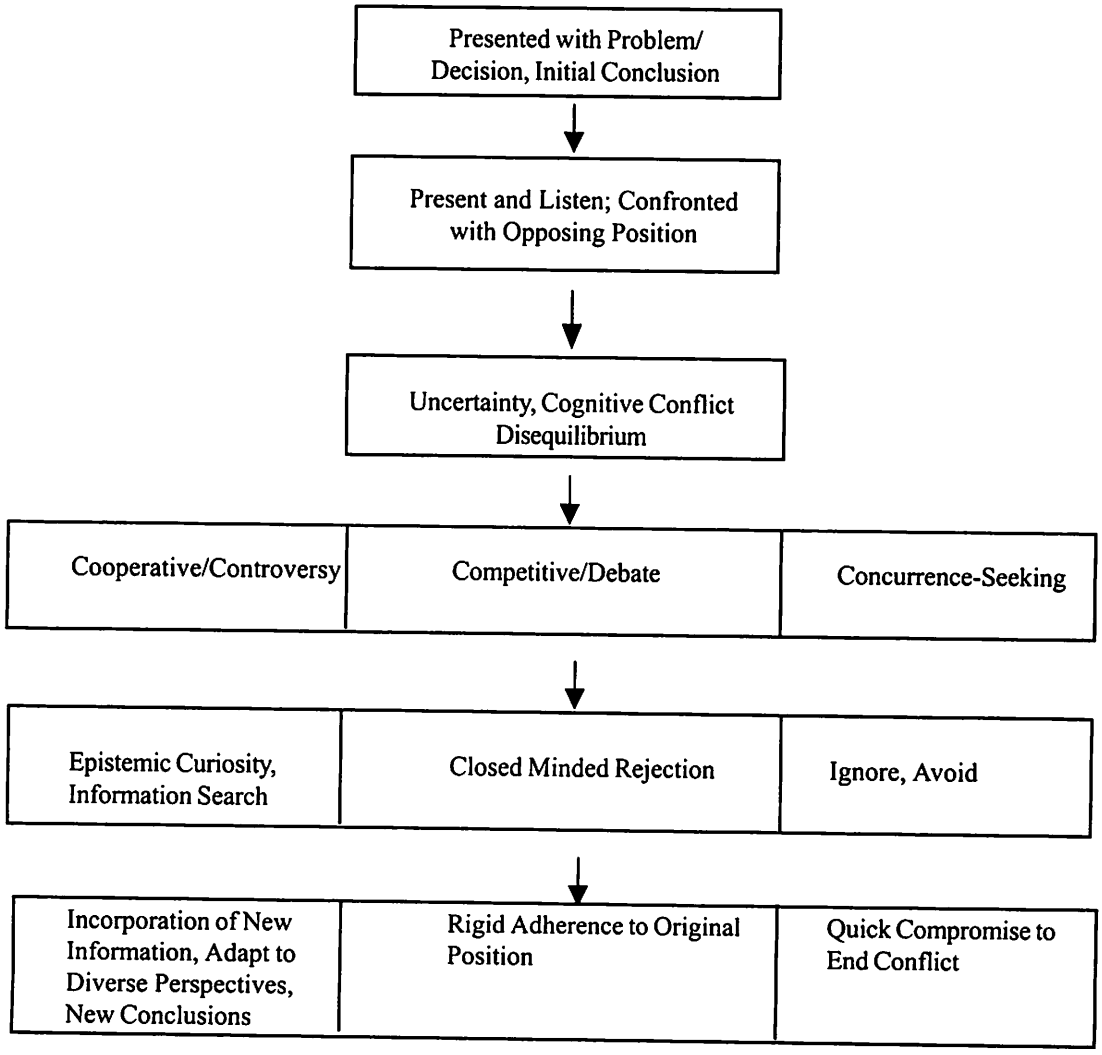

Figure 1. Processes of Controversy, Debate, Concurrence Seeking

Note. Reprinted with permission from: Johnson, D. W., & Johnson, R. (1995). *Creative controversy: Intellectual conflict in the classroom.* Edina, MN: Interaction Book Company.

in rational argument (Johnson & Johnson, 1979, 1989, 1995b, 2000).

Cooperative Goal Structure

Deutsch (1973) emphasizes that the context in which conflicts occur has important effects on whether the conflict turns out to be constructive or destructive. There are two possible contexts for controversy: cooperative and competitive. In a *cooperative situation*, individuals perceive that they can achieve their goal if and only if the other group members achieve theirs; in a *competitive situation*, individuals perceive that they can achieves their goal if and only if all others with whom they are competitively linked fail to achieve their goals. A competitive context tends to promote destructive controversy whereas a cooperative context tends to facilitate constructive controversy. Controversy within a competitive context tends to promote closed-minded disinterest and rejection of the opponent's ideas and information. Within a cooperative context, constructive controversy induces more complete and accurate communication, more accurate understanding of the opponent's position, greater utilization of others' information, greater understanding of what others are feeling and why they are feeling that way, feelings of comfort, pleasure, and helpfulness in discussing opposing positions, more open-minded listening to the opposing positions, greater motivation to hear more about the opponent's arguments, more frequently seeking out individuals with opposing opinions to test the validity of their ideas, greater trust, and the reaching of more integrated positions where both one's own and one's opponent's conclusions and reasoning are synthesized into a final position.

Skilled Disagreement

For controversies to be managed constructively, participants need collaborative and conflict-management skills (Johnson, 2003; Johnson & F. Johnson, 2003). The skills are necessary for following and internalizing these norms:

> I am critical of ideas, not people. I challenge and refute the ideas of the other participants, while confirming their competence and value as individuals. I do not indicate that I personally reject them.
>
> I separate my personal worth from criticism of my ideas.
>
> I remember that we are all in this together, sink or swim. I focus on coming to the best decision possible, not on winning.
>
> I listen to everyone's ideas, even if I don't agree.
>
> I restate what someone has said if it is not clear.
>
> I differentiate before I try to integrate. I first bring out *all* ideas and facts
>
> I try to understand both sides of the issue. I try to see the issue from the opposing perspective in order to understand the opposing position.
>
> I change my mind when the evidence clearly indicates that I should do so.
>
> I emphasize rationality in seeking the best possible answer, given the available data.
>
> I follow the *golden rule of conflict.* I act towards opponents as I would have them act toward me. I want the opposing pair to listen to me, so I listen to them. I want the opposing pair to include my ideas in their thinking, so I include their ideas in my thinking. I want the opposing pair to see the issue from my perspective, so I take their perspective.

One of the most important skills is to be able to disagree with each other's ideas while confirming each other's personal competence (Tjosvold, 1998). Disagreeing with others, and at the same time imputing that others are incompetent, tends to increase their commitment to their own ideas and their rejection of the other person's information and reasoning. Disagreeing with others while simultaneously confirming their personal competence, however, results in being better liked and in opponents being less critical of others' ideas, more interested in learning more about others' ideas, and more willing to incorporate others' information and reasoning into their own analysis of the problem. Protagonists are more likely to believe their goals are cooperative, integrate their perspectives, and reach agreement.

Another important set of skills for exchanging information and opinions within a constructive controversy is perspective-taking (Johnson, 1971; Johnson & Johnson, 1989). More information, both personal and impersonal, is disclosed when one is interacting with a person who is engaging in perspective-taking behaviors, such as paraphrasing, which communicate a desire to understand accurately. Perspective-taking ability increases one's capacity to phrase messages so that they are easily understood by others and to comprehend accurately the messages of others. Engaging in perspective-taking in conflicts results in increased understanding and retention of the opponent's information and perspective. Perspective-taking facilitates the achievement of creative, high-quality

problem solving. Finally, perspective-taking promotes more positive perceptions of the information-exchange process, of fellow group members, and of the group's work.

A third set of skills involves the cycle of differentiation of positions and their integration (Johnson & F. Johnson, 2003). Group members should ensure that there are several cycles of differentiation (bringing out differences in positions) and integration (combining several positions into one new, creative position). The potential for integration is never greater than the adequacy of the differentiation already achieved. Most controversies go through a series of differentiations and integrations before reaching a final decision.

Rational Argument

During a constructive controversy, group members have to follow the canons of rational argumentation (Johnson & Johnson, 1995b). Rational argumentation includes generating ideas, collecting relevant information, organizing it using inductive and deductive logic, and making tentative conclusions based on current understanding. Rational argumentation requires that participants keep on open mind, changing their conclusions and positions when others are persuasive and convincing in their presentation of rationale, proof, and logical reasoning.

Research Results

> He that wrestles with us strengthens our nerves, and sharpens our skill. Our antagonist is our helper.
>
> Edmund Burke, Reflection on *the Revolution in France*

The research on constructive controversy has been conducted in the last 35 years by several different researchers in a variety of settings using many different participant populations and many different tasks within an experimental and field-experimental format (see Tables 2 and 3). For a detailed listing of all the supporting studies, see Johnson and Johnson (1979, 1989, 1995b, 2000). All studies randomly assigned participants to conditions. The studies have all been published in journals (except for one dissertation), have high internal validity, and have lasted from one to sixty hours. The studies have been conducted on elementary, intermediate, and college students. Taken together, their results have considerable validity and generalizability. A recent meta-analysis provides the data to validate or disconfirm the theory. Weighted effect sizes were computed for the 28 studies included in the analyses.

Table 2

General Characteristics of Study

Characteristic	Number	Percent
1970-1979	12	43
1980-1989	16	57
Random Assigned Subjects	22	79
No Random Assignment	6	21
Grades 1 – 3	7	25
Grades 4 – 6	7	25
Grades 10 – 12	2	7
College	10	36
Adult	2	7
Published In Journals	27	96
Dissertations	1	4
1 Session	12	43
2-9 Sessions	6	21
10-20 Sessions	8	29
20+ Sessions	2	7

Note. Reprinted with permission from: Johnson, D. W., & Johnson, R. (1995). *Creative controversy: Intellectual conflict in the classroom.* Edina, MN: Interaction Book Company.

Table 3

Meta-Analysis of Academic Controversy Studies: Mean Effect Sizes

Dependable Variable	Controversy/ Concurrence Seeking	Controversy Debate Debate	Controversy Individualistic Efforts
Achievement	0.68	0.40	0.87
Cognitive Reasoning	0.62	1.35	0.90
Perspective Taking	0.91	0.22	0.86
Motivation	0.75	0.45	0.71
Attitudes Toward Task	0.58	0.81	0.64
Interpersonal Attraction	0.24	0.72	0.81
Social Support	0.32	0.92	1.52
Self-Esteem	0.39	0.51	0.85

Note. Reprinted with permission from: Johnson, D. W., & Johnson, R. (1995). Creative controversy: Intellectual conflict in the classroom. Edina, MN: Interaction Book Company.

Quality Of Decision Making And Problem Solving

In order to establish and maintain peace, individuals must be able to make effective decisions and solve problems effectively. Effective decision making and problem solving includes higher-level reasoning, accurate understanding of all perspectives, creative thinking, and openness to influence (i.e., attitude change). The results of the research indicate that compared with concurrence-seeking ($ES = 0.68$), debate ($ES = 0.40$), and individualistic efforts ($ES = 0.87$), constructive controversy tends to result in higher-quality decisions (including decisions that involve ethical dilemmas) and higher-quality solutions to complex problems for which different viewpoints can plausibly be developed. Skillful participation in a constructive controversy tends to result in (a) significantly greater ability to recall the information and reasoning contained in own and others' positions, (b) more skillfully transferring of this learning to new situations, and (c) greater generalization of principles learned to a wider variety of situations than do concurrence-seeking, debate, or individualistic efforts. The resolution of a controversy is likely to be in the direction of correct problem-solving, even when the initial conclusions of all group members are erroneous and especially when individuals are exposed to a credible minority view (as opposed to a consistent single view) even when the minority view is incorrect.

Cognitive Reasoning

When difficult issues are being discussed and effective decisions are needed to establish and maintain peace, higher-level reasoning strategies are needed. Controversy tends to promote more frequent use of higher-level reasoning strategies than do concurrence seeking ($ES = 0.62$), debate ($ES = 1.35$) or individualistic efforts ($ES = 0.90$). A number of studies on cognitive reasoning indicate that controversy is more effective than modeling and nonsocial presentation of information in influencing nonconserving children to gain the insights critical for conservation. In classrooms where students are free to dissent and are also expected to listen to different perspectives, students tend to think more critically about civic issues and be more tolerant of conflicting views.

Perspective Taking

Understanding and considering all perspectives is important if difficult issues are to be discussed, the decision is to represent the best reasoned judgment of all participants, and all participants are to help implement the decision. Constructive controversy tends to promote more accurate and complete understanding of opposing perspectives than do concurrence seeking ($ES = 0.91$), debate ($ES = 0.22$), and individualistic efforts ($ES = 0.86$). Engaging in controversy tends to result in greater understanding of another person's cognitive perspective than the absence of controversy and individuals engaged in a controversy tend to be better able subsequently to predict what line of reasoning their opponent would use in solving a future problem than were individuals who interacted without any controversy. The increased understanding of opposing perspectives tends to result from engaging in controversy (as opposed to engaging in concurrence-seeking discussions or individualistic efforts) regardless of whether one is a high-, medium-, and low-achieving student.

Creativity

High-quality decisions that subsume conflicting

positions and perspectives often require creative thought. In constructive controversies, participants tend to invent more creative solutions to problems, be more original in their thinking, generate and utilize a greater number of ideas, generate more higher quality ideas, analyze problems at a deeper level, raise more issues, have greater feelings of stimulation and enjoyment, become more emotionally involved in and committed to solving the problem, and are more satisfied with the resulting decision. Being confronted with credible alternative views, furthermore, has resulted in the generation of more novel solutions, varied strategies, and original ideas.

Attitude Change About The Issue

Open-minded consideration of all points of view is critical for deriving well reasoned decisions that integrate the best information and thought from a variety of positions. If peace is to be maintained, participants must open-mindedly believe that opposing positions are based on legitimate information and logic that, if fully understood, will lead to creative solutions that benefit everyone. Involvement in a controversy tends to result in attitude and position change. Participants in a controversy tend to reevaluate their attitudes about the issue and incorporate opponents' arguments into their own attitudes. Participating in a controversy tends to result in attitude change beyond that which occurs when individuals read about the issue and these attitude changes tend to be relatively stable over time (i.e., not merely a response to the controversy experience itself).

Motivation To Improve Understanding

Building and maintaining peace is typically enhanced by a continuing motivation to learn more about the issues being considered. Most decisions are temporary because they may be reconsidered at some future date. Continuing motivation to learn about an issue is critical for the quality of long-term decision making. Participants in a constructive controversy tend to have more continuing motivation to learn about the issue and come to the best reasoned judgment possible than do participants in concurrence seeking (ES = 75), debate (0.45), and individualistic efforts (ES = 0.64). Participants in a controversy tend to search for (a) more information and new experiences (increased specific content) and (b) a more adequate cognitive perspective and reasoning process (increased validity) in hopes of resolving the uncertainty. There is also an active interest in learning the others' positions and developing an understanding and appreciation of them. Lowry and Johnson (1981), for example, found that students involved in a controversy, compared with students involved in concurrence seeking, read more library materials, reviewed more classroom materials, more frequently watched an optional movie shown during recess, and more frequently requested information from others.

Attitudes Toward Controversy

If participants are to be committed to implement the decision and participate in future decision making, they must react favorably to the way the decision was made. Individuals involved in controversy liked the procedure better than did individuals working individualistically, and participating in a controversy consistently promoted more positive attitudes toward the experience than did participating in a debate, concurrence-seeking discussions, or individualistic decisions. Controversy experiences promoted stronger beliefs that controversy is valid and valuable. The more positive the attitudes toward the process of making the decision, the more committed participants may feel to implement the decision.

Attitudes Toward Decision Making

If participants are to be committed to implement the decision and participate in future decision making, they must consider the decision worth making. Individuals who engaged in controversies tended to like the decision making task better than did individuals who engaged in concurrence-seeking discussions (ES = 0.63).

Interpersonal Attraction And Support Among Participants

Decision making, to be effective, must be conducted in ways that bring individuals together, not create ill-will and divisiveness. Within controversy there is disagreement, argumentation, and rebuttal that could create difficulties in establishing good relationships. Constructive controversy, however, has been found to promote greater liking among participants than did debate (ES = 0.72), concurrence-seeking (ES = 0.24), or individualistic efforts (ES = 0.81). Debate tended to promote greater interpersonal attraction among participants than did individualistic efforts (ES = 0.46). In addition, constructive controversy tends to promote greater social support among participants than does debate (ES = 0.92), concurrence-seeking (ES = 0.32), or individualistic efforts (ES = 1.52). Debate tended to promote greater social support among participants than did individualistic efforts (ES = 0.85).

Self-Esteem

Participation in future decision making is enhanced when participants feel good about themselves as a result of helping make the current decision, whether or not they agree with it. Constructive controversy tends to promote higher self-esteem than does concurrence-seeking (ES = 0.39), debate (ES = 0.51), or individualistic efforts (ES = 0.85). Debate tends to promote higher self-esteem than does individualistic efforts (ES = 0.45).

Structuring Constructive Controversies

> Conflict is the gadfly of thought. It stirs us to observation and memory. It instigates invention. It shocks us out of sheeplike passivity, and sets us at noting and contriving...Conflict is a "sine qua non" of reflection and ingenuity.
>
> John Dewey, *Human Nature and Conduct: Morals Are Human*

In order to use constructive controversy to foster academic learning, you implement the following procedure (Johnson & R. Johnson, 1979, 1989, 1995b, 2000): The task must be structured (a) cooperatively and (b) so that there are at least two well-documented positions (pro and con). Students are typically randomly assigned to groups of four. Each group is then divided into two pairs. The pairs are randomly assigned to represent either the pro or con position. The instructional materials are prepared so that group members know what position they have been assigned and where they can find supporting information. Teachers may wish to help students "*get in role*" by presenting the issue to be decided in an interesting and dramatic way. Teachers structure positive interdependence by assigning two group goals of producing a group report detailing the nature of the group's decision and its rationale and individually taking a test on both positions. The purpose of the constructive controversy is to maximize each student's learning. The constructive controversy procedure is as follows:

1. *Research, Learn, and Prepare Position.* Students are randomly assigned to groups of four, each of which is divided into two pairs. One pair is assigned the pro position and the other pair is assigned to the con position. Each pair is to prepare the best case possible for its assigned position by:

   a. *Researching the assigned position and learning all relevant information.* Students are to read the supporting materials and find new information to support their position. The opposing pair is given any information students find that supports its position.

   b. *Organizing the information into a persuasive argument* that contains a thesis statement or claim ("George Washington was a more effective President than Abraham Lincoln"), the rationale supporting the thesis ("He accomplished a, b, and c"), and a logical conclusion that is the same as the thesis ("Therefore, George Washington was a more effective President than Abraham Lincoln").

   c. *Planning how to advocate the assigned position* effectively to ensure it receives a fair and complete hearing. Make sure both pair members are ready to present the assigned position so persuasively that the opposing participants will comprehend and learn the information and, of course, agree that the position is valid and correct.

2. *Present and Advocate Position.* Students present the best case for their assigned position to ensure it gets a fair and complete hearing. They need to be forceful, persuasive, and convincing in doing so. Ideally, more that one media will be used. Students are to listen carefully to and learn the opposing position, taking notes and clarify anything they do not understand.

3. *Engage In An Open Discussion In Which There is Spirited Disagreement.* Students discuss the issue by freely exchanging information and ideas. Students are to (a) argue forcefully and persuasively for their position (presenting as many facts as they can to support their point of view), (b) critically analyze the evidence and reasoning supporting the opposing position, asking for data to support assertions, (c) refuting the opposing position by pointing out the inadequacies in the information and reasoning, and (d) rebutting attacks on their position and presenting counter arguments. Students are to take careful notes on and thoroughly learn the opposing position. Students are to give the other position a "trial by fire" while following the norms for constructive controversy. Sometimes a "time-out" period will be provided so students can caucus with their partners and prepare new arguments. The teacher may encourage more spirited arguing, take sides when a pair is in trouble, play devil's advocate, ask one group to observe another group engaging in a spirited argument, and generally stir up the discussion.

4. *Reverse Perspectives.* Students reverse perspectives and present the best case for the opposing position. Teachers may wish to have students change chairs. In presenting the opposing position sincerely and forcefully (as if it was yours), students may use their notes and add any new facts they know of. Students should strive to see the issue from both perspectives simultaneously.

5. *Synthesize.* Students are to drop all advocacy and find a synthesis on which all members can agree. Students summarize the best evidence and reasoning from both sides and integrate it into a joint position that is a new and unique. Students are to:

   a. Write a group report on the group's synthesis with the supporting evidence and rationale. All group members sign the report indicating that

they agree with it, can explain its content, and consider it ready to be evaluated. Each member must be able to present the report to the entire class.

b. Take a test on both positions. If all members score above the preset criteria of excellence, each receives five bonus points.

c. Process how well the group functioned and how its performance may be improved during the next constructive controversy. The specific conflict management skills required for constructive controversy may be highlighted.

d. Celebrate the group's success and the hard work of each member to make every step of the constructive controversy procedure effective.

While the groups engage in the constructive controversy procedure, teachers monitor the learning groups and intervene to improve students' skills in engaging in each step of the constructive controversy procedure and use the social skills appropriately. At the end of each instructional unit, teachers evaluate students' learning and give feedback. The learning groups also process how well they functioned by describing what member actions were helpful (and unhelpful) in completing each step of the constructive controversy procedure and make decisions about what behaviors to continue or change.

Constructive Controversy And Democracy

Thomas Jefferson and his fellow revolutionaries believed that free and open discussion should serve as the basis of influence within society, not the social rank within which a person was born. American democracy was, therefore, founded on the premise that "truth" will result from free and open-minded discussion in which opposing points of view are advocated and vigorously argued. Every citizen is given the opportunity to advocate for his or her ideas and to listen respectfully to opposing points of view. Once a decision is made, the minority is expected to go along willingly with the majority because they know they have been given a fair and complete hearing. To be a citizen in our democracy, individuals need to internalize the norms for constructive controversy as well as mastering the process of researching an issue, organizing their conclusions, advocating their views, challenging opposing positions, making a decision, and committing themselves to implement the decision made (regardless of whether one initially favored the alternative adopted or not). In essence, the use of constructive controversy teaches the participants to be active citizens of a democracy.

Intervening To Create Peace

Intervening to create peace can involve four different levels of intervening (Duckitt, 1992; Oscamp, 2000): Interventions may focus on (a) changing human genetics and evolution, (b) instigating new laws and widespread norms through organizations such as the United Nations and the World Court as well as through international treaties, (c) influencing human behavior through education, mass media, work roles and group and interpersonal processes, and (d) changing personality characteristics through such processes as psychotherapy. Of these four levels, Gandhi believed that the use of Level 3 interventions such as education to influence children is the most feasible and effective approach. He believed that education is our continual hope, as at any point in time we can change subsequent generations and improve the human condition. Our progress towards increased humanity and peace may be education based.

Creating A Tipping Point

A *tipping point* is a dramatic moment in which everything changes all at once (Gladwell, 2000). It is usually discussed in terms of epidemics and happens when three influences converge: contagion (the behavior and beliefs of a few people spread to lots of other people), many small incremental influences converge in the same direction, and dramatic change (not gradual) occurs (i.e., there is a sudden geometric progression in the behavior or beliefs). Gladwell (2000) presents an example of an epidemic of syphilis that occurred in Baltimore in the mid-1990s. Previously, for years and years the number of children born with syphilis was low but stable. Then suddenly, in 1995-1996, the number increased by 500 percent. When you look at this on a graph, a flat line runs straight for many years and then suddenly rises almost at a right angle. Three explanations were given for the sudden rise: (a) crack cocaine brought more people into poor neighborhoods to buy drugs and changed the patterns of social connections in the city, (b) there was a breakdown in medical services in the city's poorest neighborhoods, and (c) urban renewal destroyed the neighborhood where syphilis was concentrated and relocated the people throughout the city. Acting together, these (and perhaps other) unknown changes created a tipping point. The same dynamics could happen with peace. Through education it is possible to give every student 12 years or more years of training and practice in how to engage in constructive controversy so that the difficult issues that form the basis for destructive conflict and potential violence may be discussed and understood, the cooperative context required for conflicts to be managed constructively may be established and maintained, and the civic values underlying constructive controversy and cooperation. Doing so will result in adults who are more competent in establishing and

maintaining peace and a point will be reached in which societies "tip" towards peace.

Influences Toward Tipping Points In Education

When faced with problems such as creating peace, humans typically turn to education. H. G. Wells (1927), for example, in reflecting on World War I and the current state of the world, pointed out that human beings are embarked upon a race between education and catastrophe. Lorenz (1963), in reflecting on World War II and the subsequent cold war and nuclear arms race, noted that an unprejudiced observer from another planet, observing humans as with their combination of nuclear weapons and aggressiveness, would not prophesy long life for the species. He believes that humans are at a crossroads at which they either educate themselves in how to manage conflicts constructively or they continue along a road leading to the extinction of the human species. He notes that in the evolutionary view, this would be about as significant as the extinction of the ichthyosaur.

Wells, Lorenz, and many other social scientists, therefore, believe that the hope for peace lies in implementing peace education in schools and universities. The history on innovation in education, however, indicates that not every innovation is equal (Johnson & Johnson, 1999). Many new practices in schools are quickly and even widely adopted, and then dropped quickly and disappear over night. The innovations that tend to last are characterized by being based on clear procedures that are based on a theory that has considerable validating research. The implementation of the procedures then reveal inadequacies in the theory which result in revising the theory, conducting new research to validate it, and modifying the operational procedures accordingly. This interaction among theory, research, and practice is perhaps the most powerful guarantee that an innovation will be adopted in schools and universities, institutionalized, and maintained over time, thus contributing to a tipping point in a society and the world. Constructive controversy (Johnson & Johnson, 1995b), as well as the related practices of cooperative learning (Johnson & Johnson, 1999) and the Teaching Students To Be Peacemakers Program (Johnson & Johnson, 1995a, 1996a, 2000) procedures, are derived from theory, have been validated by research, and been implemented in schools throughout the world.

Summary

Peace is freedom from war or strife. In order to understand peace, it is also necessary to understand war, cooperation, and conflict. Peace education has been defined in numerous ways, but includes (a) facing the difficult issues that must be discussed in order for peace to be established and maintained, (b) establishing a procedure that all parties agree to use to discuss the issues involved in the conflict, (c) training students how to use the procedure skillfully, and (d) incorporating the use of the procedure into students' personal identity so that it will be habitually used.

Establishing peace requires discussing difficult issues (often involving ethnic, cultural, or religious differences) on which agreement may not be possible, but some accommodation must be worked out. Doing so is not easy, as often the discussions are seen as opportunities for advocacy with little interest in learning anything about other points of view. Those who do not have direct personal experience are seen as lacking expertise and having little to contribute. In order to discuss these difficult issues a procedure is needed that allows constructive discussions to take place. One such procedure is constructive controversy.

In well structured controversies, participants make an initial judgment, present their conclusions to other group members, are challenged with opposing views, become uncertain about the correctness of their views, actively search for new information and understanding, incorporate others' perspectives and reasoning into their thinking, and reach a new set of conclusions. This process results in significant increases in the quality of decision making and problem solving (including higher-levels of cognitive and moral reasoning, perspective taking, creativity, and attitude change about the issue), motivation to learn more about the issue, positive attitudes toward the controversy and decision making processes, the quality of relationships, and self-esteem. While the constructive controversy process can occur naturally, it may be consciously structured in decision making and learning situations. This involves dividing a cooperative group into two pairs and assigning them opposing positions. The pairs then (a) develop their position, (b) present it to the other pair and listen to the opposing position, (c) engage in a discussion in which they attempt to refute the other side and rebut attacks on their position, (d) reverse perspectives and present the other position, and (e) drop all advocacy and seek a synthesis that takes both perspectives and positions into account. Engaging in the constructive controversy procedure skillfully provides an example of how conflict creates positive outcomes.

The educational use of academic controversy may be utilized in any grade level and in any subject matter. Engaging in the controversy process needs to pervade school life so that students develop considerable expertise in its use and incorporate the process into their identity. Any time students participate in the controversy procedure, they are getting a lesson in peace education. They are also getting a lesson in democracy. The academic controversy procedure is a Level 3 intervention aimed at creating a tipping point in which enough individuals become skillful in the constructive controversy procedure that they will use it to establish and maintain peace. The possibility of this taking place is strengthened by the foundation of theory and research on which the controversy procedure is based.

References

Deutsch, M. (1973). *The resolution of conflict.* New Haven, CT: Yale University Press.

Duckitt, J. (1992). *The social psychology of prejudice.* New York: Praeger.

Gladwell, M. (2000). *The tipping point.* Boston: Little, Brown, & Company.

Janis, I. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes.* Boston, MA: Houghton-Mifflin.

Johnson, D. W. (1970). *The social psychology of education.* New York: Holt, Rinehart, & Winston.

Johnson, D. W. (1971). Role reversal: A summary and review of the research, *International Journal of Group Tensions, 1*, 318-334.

Johnson, D. W. (2003). *Reaching out: Interpersonal effectiveness and self-actualization* (8th ed.). Boston: Allyn & Bacon.

Johnson D. W., & Johnson, F. (2003). *Joining together: Group theory and group skills* (8th ed.). Boston: Allyn & Bacon.

Johnson, D. W., & Johnson, R. (1979). Conflict in the classroom: Constructive controversy and learning. *Review of Educational Research, 49,* 51-61.

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. (1995a). *Teaching students to be peacemakers.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. (1995b). *Creative Constructive controversy: Intellectual challenge in the classroom.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R.. (2000). Civil political discourse in a democracy: The contribution of psychology. *Peace and Conflict: Journal of Peace Psychology, 6*(4), 291-317.

Johnson, D. W., Johnson, R., & Holubec, E. (2002). *Circles of learning: Cooperation in the classroom* (8th ed.). Edina, MN: Interaction Book Company.

Lorenz, C. (1963). *On aggression.* New York: Hartcourt, Brace, and World.

Oscamp, S. (2000). Multiple paths to reducing prejudice and discrimination. In S. Oscamp (Ed.), *Reducing prejudice and discrimination* (pp. 1-19) Mahwah, NJ: Lawrence Erlbaum.

Tjosvold, D. (1998). Cooperative and competitive goal approach to conflict: Accomplishments and challenges. *Applied Psychology: An International Review, 47*(3), 285-342.

Tjosvold, D., & Johnson, D. W. (Eds.). (1983). *Productive Conflict Management: Perspectives for Organizations.* New York: Irvington.

Trosset, C. (1998). Obstacles to open discussion and critical thinking: The Grinnell College study. *Change, 30*(5), 44-49.

Wells, H. G. (1927). *Outline of history.* New York: Macmillan Press.

David W. Johnson is Professor of Educational Psychology at the University of Minnesota. He specializes in cooperation and competition, conflict resolution, and organizational change.

Roger T. Johnson is a Professor of Education at the University of Minnesota. He specializes in science education.

Research Tools

Editor: David M. Shannon
Auburn University

We are pleased to offer another section pertaining to methodological issues. This section provides an opportunity for the readership to learn about and generate a discussion about a variety of issues as they pertain to educational researchers. The current section includes five articles that address issues facing educational researchers as they manage, organize, analyze, and summarize data. This section includes articles that confront the critical issues of reliability and validity.

The first two articles address the issue of data analysis. Although the creation and refinement of statistical software (e.g. SAS, SPSS) has made sophisticated analysis more readily available to everyone, many research management, organization, and analytical tasks can be accomplished using commonly available, and more affordable, software. "Using Common Software to Aid Research Tasks," written by Packard, illustrates the use of common software (i.e., Excel and Word) to assist you with managing, sorting, analyzing, and visually summarizing quantitative and qualitative data. The second article, "Show Me the Numbers: Learning and Using Quantitative Skills to Improve Education Leadership," written by Schiller, Dragone, & Trapini provides an overview of the application of common software and statistical analysis tools in dealing with real educational issues in a graduate course.

The remaining two articles directly address the integral concepts of reliability and validity. Providing supportive evidence of these properties is essential for published research. Unfortunately, reviews across multiple journals from counseling psychology, psychology, and education have not been supportive, concluding that such properties are not evaluated sufficiently, reported in less than 50% of published articles (Kieffer, Reese, & Thompson, 2001; Meier & Davis, 1990; Thompson & Synder, 1998; Vacha-Hasse, Ness, Nilsson, & Rheetz, 1999). In addition, when reliability information from a published instrument is reported, it is often based on results from a previous sample without demonstrating the comparability of the current sample to that upon which the psychometric properties were estimated (Whittingham, 1998). The sufficiency of information reported by authors developing their own measures is even less encouraging. An editorial in the September 2003 issue of the Journal of Educational Psychology reinforces the need to address these concerns (Harris, 2003). For these reasons, reliability generalization studies have become more prominent in the field (see Vacha-Haase, 1998) to determine the stability of reliability of widely used instruments across different samples.

"Reliability of Scores and the Researcher," written by Forbes and Ross, addresses the foundational concepts of classical test theory, illustrating the uses of commonly used reliability indices and providing practical guidelines for their use. Finally, "Learning Behaviors in a College Setting: A Psychometric Examination of the LASSI," written by Schaefer and Worrell, offers an extensive examination of the psychometric properties of the LASSI. This study investigates these properties in terms of internal consistency reliability, predictive validity, and construct validity.

We hope this section generates discussion among the readership and generates future submissions for consideration in this section. If you would like to submit a manuscript for this section, or simply discuss some ideas, please contact:

David M. Shannon
Professor
Auburn University
4036 Haley Center - EFLT
Auburn University, AL 36849-5221
(334) 844-3071
shanndm@auburn.edu

Harris, K. R. (2003). Editorial: Is the Work as Good as It Could Be? *Journal of Educational Psychology, 95*(3), 451-452.

Kieffer, K.M, Reese, R.J. & Thompson, B. (2001). Statistical techniques employed in AERJ and JCP articles from 1988 to 1997: A methodological review. *Journal of Experimental Education, 69*(3), 280-309.

Meier, S. T. & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology, 37*, 113-115.

Vacha-Hasse, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*, 335-341.

Vacha-Hasse, T. (1998). Reliability generalization: Exploring variance in measurement error effecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20.

Whittington, D. (1998). How well do researchers report their measures? An evaluation of measurement in published educational research. *Educational and Psychological Measurement, 56*, 21-37.

Using Common Software to Aid in Research Tasks

Abbot Packard

State University of West Georgia

Often faculty and students do not have daily access to software created to handle research tasks. But most of these individuals have access to common software built into suites such as MS Office. It is the intention of this article to present functions to help the researcher with either an intermediate step to finish a project. Using two products of MS Office© Excel© and Word©, I will illustrate how these two products can aid researchers with quantitative and qualitative research. Word will be discussed in context of qualitative research. Excel© contains many functions which will allow both quantitative and qualitative research such as simple analysis (math or statistical), graphics, databases and qualitative analysis.

Excel's Basic Functions

Excel is a commonly available spreadsheet software program that is frequently used to organize, summarize, and report data for a wide variety of purposes (e.g. budgeting, grading). Excel makes data entry very easy as it has options to move in any of four directions after entry of each piece of data.

Tool Bar					
	Options				
		Edit			
			Move selection after enter		
				Direction	
					Down Right Up Left

Figure 1. Steps to alter data entry in Excel

Data Manipulation. Manipulation such as combining scores, creating subsets scores, coding of variables and labeling can be accomplished in Excel© prior to importing into a more sophisticated statistical analysis package such as SAS or SPSS. Manipulations which are common in data analysis can be easily performed in Excel. Combining scores on a survey or test, whether it is total scores which are needed or subscale scores, can be performed without designing a formula. Descriptive statistics which are available with a few clicks of the mouse include: Mean, standard error, median, mode, standard deviation. Kurtosis, Skewness, range and more. Random number generation (a worthwhile tool in determining a sample) can also be done using many difference generation methods from uniform through discrete. An Example is the Auto Sum function built into the toolbar.

Tool Bar	
Σ	Auto Sum

Figure 2. Summation symbol found on the tool bar of Excel © or in the functions

Creating and applying formulas. There are 80 statistical and 63 mathematical formulas which will accomplish simple analyses by using the built in add one in Excel.

Tool Bar			
	f_x	categories	
			Math & Trig and Statistical

Figure 3. Steps to find the functions in Math and Statistics

Researchers who use formulas can create them within the Excel program to perform repeated analysis. Whether simple or complex, once created, they can be saved and used when needed. For example, effect sizes are strongly recommended or required by many journals. When not directly discussed in an article, they can easily be created from the reported mean(s) and standard deviation(s) using a few simple Excel functions to create the appropriate formula. So in addition to the standard offerings by Excel more can be created adding to the variety of analyses available. Once a formula is created it can be saved to be used multiple times.

Creating Graphs. Graphic capabilities of Excel help display results effectively. There are 14 different types of graphs with multiple variations of each. These graphs are created following a step by step procedure integrated within Excel. Visual displays of data make results from research studies clearer to the reader, especially when multiple pieces of information are summarized. Visual displays tends to capture the big picture and help focus the discussion of results on the most critical findings.

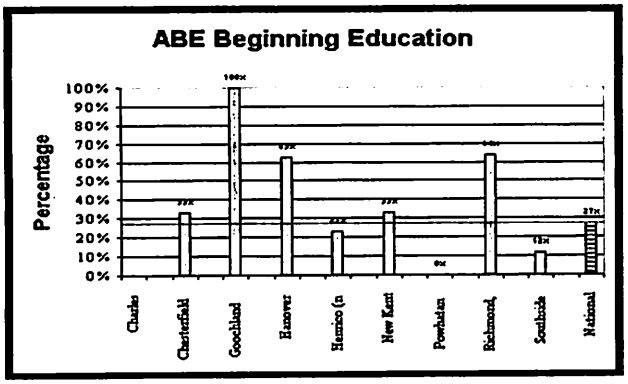

Figure 4. An example of a Column graph illustrating categorical data.

When creating figures for APA style publications there are certain guidelines required (see Figure 5). APA style dictates that the top border and right axis are removed. Lettering is professional using a combination of caps and lower case running parallel to axis. Data lines are drawn with smooth curves and units of measure are specified. All this can easily be accomplished in Excel.

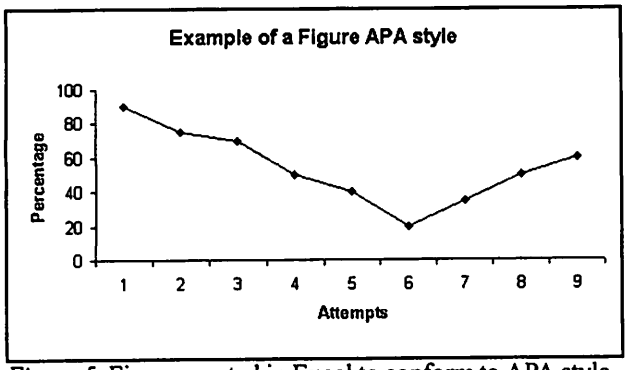

Figure 5. Figure created in Excel to conform to APA style.

Figures such as graphs illustrate the data to enable the reader to comprehend what the numbers are suggesting. There are different types of graphs which are commonly used to convey the results or findings. *Scatter plots* present a picture of a relationship between variables and may demonstrate a correlation. These graphs can be combined with a line to indicate a best fit linear relationship. A *line graph* should be used with independent and dependent quantitative variables such as shown in Figure 5 above. A relationship between the two variables can be easily seen by the reader. *Column (Bar) graphs* are often used to summarize categorical data. Each bar represents a category with the Y axis measuring the data or number for each category. *Circle graphs*(pie charts)are used to demonstrate the percentages and portion of variables included in the study. Caution should be used when applying a circle or pie graphs as they are not very clear when using more than five different conditions. Tables reporting results can also be formatted easily in Excel or Word but neither program completes the formatting automatically for APA.

Database functions. Creation of a database for organization of references can be accomplished using Excel or Word. This takes several steps but will save time when writing a paper. As you read an article, you make a judgment of its usefulness for a work in progress or a future work. Upon seeing something which illustrates a point or argument well, it can be noted and saved for later use. Understanding what location in an article these points might be used allows a coding process to be useful such as: (1) the introduction, (2) literature review and (3) conclusion. By coding these phrases, they can be easily retrieved later when writing simply by sorting by the location. Building the database can include as many options as you prefer. Suggested column headers may include the following: Coding of location (where it might be used in the article), author, year, title, volume, issue, page numbers and notes. (See Figure 6)

Notes can be stored by selecting word wrap found under *Format, Cells, Alignment*, and then check box of *Wrap Text.* Once references are stored and the writing begins, the location code can be sorted and the references to support are easily found. A secondary sort could be used to sort by year to assist in creating an historical order to the writing. The database is easy to create and use and available to most.

location	author	yr	title	Journal	volume	issue	page	notes
1	Ackerman, P.L.	1988	Determinants of Individual differences during skill acquisition: Cognitive Abilities and Informational processing	Journal of Experimental Psychology: General	117	3	288-318	Acquisition on with ability determinants of individual differences in...

Figure 6. Example of a database created for references

Paper copies of references can be stored in inexpensive storage boxes in some order (alphabetical) for easy retrieval during the writing process.

Qualitative applications. Qualitative research can be aided with both Excel and Word. Narrative data can be converted to a table or into a spreadsheet. Each row can hold a sentence or paragraph (optional to the researcher) which can be read and coded. By adding a column to the existing table, the coding can be stored and then later sorted according to the researcher's preferences. By adding another column a sub code can be added and that can be sorted as well. In this way the researcher is aided by a method that helps organize his/her coded themes and allows him/here to support such themes with exact quotes from the narrative. Some qualitative researchers prefer simple line numbering leaving the narrative in the same context as it was recorded. This allows the researcher to precisely locate the origin of all data and the source of coded themes. This is done by selecting the entire text and then following the steps of line numbering.

File				
	Page setup			
		Layout		
			Line Numbers	Add line numbers

Figure 6. Steps to add line numbers to your document in Word.

During collection of data whether by interview, focus group or historical documents, many different words may be used for the same context. For example, content such as student, might be referred to as children, boys, girls, kids, etc. By transcribing the data a researcher may begin to see these words as being similar. There are other options to make the task easier by counting the frequency of words used in the transcript. One such option is a web site www.georgetown.edu/cball/webtools/web_freqs.html which will count the frequency of all words in the document. Then words with similar meaning can be coded for later analysis. Word has a function called replace found under the Edit on the tool bar. It will find words and replace them with what the researcher chooses. It is perhaps better to replace the word with the same word and a bracketed key word such as *kids [student]*. This method allows the original word to remain with the new capability to find the key word.

Word also allows the researcher to add comments. Depending on the version you have, Word will create the comments in different display modes. Word XP performs the task with the following results: Comment: This will illustrate. Qualitative analysis is complex at best and the use of Word is not the final answer. For researchers who analyze qualitative data consistently, programs such as Nudist and Atlas might be more appropriate. The analysis proceeds in similar fashion and part time qualitative researchers can use the capabilities of a word processor.

Transcribing the data is an initial step regardless of how you proceed with the analysis. The second step is to build a coding system that requires you to have knowledge of the material (content). The third step is highlighting or coding the text. This step can also be performed with word. The possibilities are not limited to the few mentioned here but these provide the researcher with some tools to work with using common software. There are many sources for these and other solutions to use common software for analysis. One of the best sources, however, is to listen to our peers and find different tricks to save time and make task easier.

Resource List

Dretzke, B. J. (2001). *Statistics with Excel* (2nd edition). Upper Saddle River, NJ: Prentice Hall

Halvorson, M. & Young, M.J. (2001).*Microsoft Office XP Inside Out*. Buffalo, NY: Microsoft Press.

Lehmann, M. & Zeitz, P. (1997). *Statistical Exploration with Microsoft Excel*. Pacific Grove, CA Duxbury Press.

Microsoft Office Online. Retrieved September 14. 2003, from http://office.microsoft.com/assistance/

Sincich, T., Levine, D., Stephen, D. & Levine, D. M. (1999). *Practical Statistics by Example using Microsoft Excel*. Upper Saddle River, NJ: Prentice Hall.

Walkenbach, J. (2002. *Excel Charts*. Hoboken, NJ :John Wiley & Sons.

Abbott L. Packard is Assistant Professor of Educational Leadership and Professional Studies at the State University of West Georgia.

Show Me the Numbers:
Learning and Using Quantitative Skills to Improve Education Leadership

Kathryn S. Schiller
State University of New York, University at Albany

Joseph P. Dragone
Cohoes City School District

Annette Trapini
State University of New York, University at Albany

In this era of information technology and increasing accountability, education leaders are expected to use the vast amount of data collected about students, faculty, and schools in planning and evaluation of educational programs at all levels. They are required to be not only knowledgeable consumers of statistical reports, but also analysts

Paper copies of references can be stored in inexpensive storage boxes in some order (alphabetical) for easy retrieval during the writing process.

Qualitative applications. Qualitative research can be aided with both Excel and Word. Narrative data can be converted to a table or into a spreadsheet. Each row can hold a sentence or paragraph (optional to the researcher) which can be read and coded. By adding a column to the existing table, the coding can be stored and then later sorted according to the researcher's preferences. By adding another column a sub code can be added and that can be sorted as well. In this way the researcher is aided by a method that helps organize his/her coded themes and allows him/here to support such themes with exact quotes from the narrative. Some qualitative researchers prefer simple line numbering leaving the narrative in the same context as it was recorded. This allows the researcher to precisely locate the origin of all data and the source of coded themes. This is done by selecting the entire text and then following the steps of line numbering.

File				
	Page setup			
		Layout		
			Line Numbers	Add line numbers

Figure 6. Steps to add line numbers to your document in Word.

During collection of data whether by interview, focus group or historical documents, many different words may be used for the same context. For example, content such as student, might be referred to as children, boys, girls, kids, etc. By transcribing the data a researcher may begin to see these words as being similar. There are other options to make the task easier by counting the frequency of words used in the transcript. One such option is a web site www.georgetown.edu/cball/webtools/web_freqs.html which will count the frequency of all words in the document. Then words with similar meaning can be coded for later analysis. Word has a function called replace found under the Edit on the tool bar. It will find words and replace them with what the researcher chooses. It is perhaps better to replace the word with the same word and a bracketed key word such as *kids [student]*. This method allows the original word to remain with the new capability to find the key word.

Word also allows the researcher to add comments. Depending on the version you have, Word will create the comments in different display modes. Word XP performs the task with the following results: Comment: This will illustrate. Qualitative analysis is complex at best and the use of Word is not the final answer. For researchers who analyze qualitative data consistently, programs such as Nudist and Atlas might be more appropriate. The analysis proceeds in similar fashion and part time qualitative researchers can use the capabilities of a word processor.

Transcribing the data is an initial step regardless of how you proceed with the analysis. The second step is to build a coding system that requires you to have knowledge of the material (content). The third step is highlighting or coding the text. This step can also be performed with word. The possibilities are not limited to the few mentioned here but these provide the researcher with some tools to work with using common software. There are many sources for these and other solutions to use common software for analysis. One of the best sources, however, is to listen to our peers and find different tricks to save time and make task easier.

Resource List

Dretzke, B. J. (2001). *Statistics with Excel* (2nd edition). Upper Saddle River, NJ: Prentice Hall

Halvorson, M. & Young, M.J. (2001).*Microsoft Office XP Inside Out*. Buffalo, NY: Microsoft Press.

Lehmann, M. & Zeitz, P. (1997). *Statistical Exploration with Microsoft Excel*. Pacific Grove, CA Duxbury Press.

Microsoft Office Online. Retrieved September 14. 2003, from http://office.microsoft.com/assistance/

Sincich, T., Levine, D., Stephen, D. & Levine, D. M. (1999). *Practical Statistics by Example using Microsoft Excel*. Upper Saddle River, NJ: Prentice Hall.

Walkenbach, J. (2002. *Excel Charts*. Hoboken, NJ :John Wiley & Sons.

Abbott L. Packard is Assistant Professor of Educational Leadership and Professional Studies at the State University of West Georgia.

Show Me the Numbers:
Learning and Using Quantitative Skills to Improve Education Leadership

Kathryn S. Schiller
State University of New York, University at Albany

Joseph P. Dragone
Cohoes City School District

Annette Trapini
State University of New York, University at Albany

In this era of information technology and increasing accountability, education leaders are expected to use the vast amount of data collected about students, faculty, and schools in planning and evaluation of educational programs at all levels. They are required to be not only knowledgeable consumers of statistical reports, but also analysts

producing reports accessible to diverse audiences of fellow educators, policy makers, parents, and the public. In this article, we discuss how rigorous instruction on statistical analysis can be balanced with practical experience working with data relevant to school administration and improvement. This approach to teaching applied statistics prepares future education leaders to effectively use quantitative data for problem solving, as illustrated by examples of analyses motivated by school improvement efforts by practicing education leaders.

With demands for higher academic standards and greater school accountability, school leaders are increasingly called upon to become more sophisticated researcher-practitioners. The *Leave No Child Left Behind Act of 2001* required that states develop curriculum standards and start testing all students in grades 3-5, grades 6-9, and grades 10-12 on their progress toward those standards (No Child Left Behind, 2001). In addition to overall progress toward meeting state standards, the law called for monitoring within each school the progress of students who are economically disadvantaged, are members of racial or ethnic minority groups, have disabilities, or have limited English proficiency. Schools whose students fail to make state-defined adequate progress toward meeting the state standards will be subjected to increasingly severe sanctions over five years culminating with restructuring, such as state takeover or conversion to a charter school.

New York is among those states that since the mid-1990s have been adopting similar reforms linking student performance to consequences for schools and encouraging the use of data in developing school improvement plans. Current state policy requires testing all 4th and 8th graders in English and mathematics, as well as requiring students pass New York State Regents' examinations in English, mathematics, history, and biology in order to graduate from high school (New York State Department of Education, n.d.). The lowest performing schools and those failing to make adequate progress toward the state's performance standards are subject to state review and potential reorganization. All districts are encouraged by the state to develop comprehensive or strategic district improvement plans based on empirical data, set goals related to increasing student achievement, and specify procedures for monitoring steps toward those goals. These efforts require both district and school leaders to undertake more meaningful analyses and interpretations of data concerning students and teachers than in the past to diagnose potential problems and evaluate interventions.

To meet this increasing demand for data-driven decisions, programs preparing future educational leaders need to provide their students with a working knowledge of applied statistics and experience undertaking analyses pertinent to their schools (Creighton 2001b, McNamara 2000). In this article we outline the three general design principals used when developing an applied quantitative data analysis course (EAPS614) for the Department of Educational Administration and Policy Studies at the University at Albany. First, the course should be accessible to all students, even those with only basic mathematical or computer skills. Second, the course should focus on essential concepts and skills that build upon each other in a coherent manner to help students understand connections between them. Third, the course should be structured to integrate theory and practice using, where possible, real world examples and stressing substantive interpretations. The success of this approach is illustrated with several analyses of real school data undertaken as school leaders by former students.

Training Future Leaders in Data Analysis

Many students in educational leadership programs lack confidence in their mathematical skills and, thus, tend to approach courses perceived as involving quantitative reasoning such as statistics with great apprehension (Creighton 2001a). This anxiety about mathematics is not unusual among adult students in technical and professional preparation programs but can impact their performance in statistics courses (Fenster 1992). Very few students in graduate programs, however, actually lack the understanding of elementary algebra and basic quantitative reasoning skills that are needed to gain a working knowledge of basic statistical analysis. Instead, the major roadblocks students encounter are their fear of failure, inexperience verbalizing quantitative concepts, a lack of motivation, and a regression to dualism (McMillan 2001). McMillan (2001) recommends that effective statistics courses motivate students, create a positive learning environment, focus on deep understanding of essential concepts, and effectively use technology. McNamara (2000) also recommends illustrating statistical concepts using school-related, real-world data and problems. Using strategies similar to those outlined by McMillan (2001) and McNamara (2000), educational leadership programs can help all students become confident and proficient data analysts through courses that feature carefully articulated curricula and assignments focused on key concepts and basic processes grounded in examples relevant to educational leadership.

Focusing on the motto "Don't panic," EAPS614 is a semester-long, 3-credit course designed to provide students with a working knowledge of quantitative analysis based on an understanding of statistics' theoretical foundations, experience applying basic statistical procedures, and good data analysis habits. Wide in scope for an introductory course, EAPS614 covers topics from types of variables and descriptive statistics to Chi-square analyses and regression. Students use Microsoft EXCEL® to analyze data sets for their four homework assignments and a group project. In addition to the lectures, laboratory sessions and teaching assistants are also available to students seeking additional help mastering the material. Students' knowledge of statistical concepts and procedures, in addition to their ability

to interpret EXCEL® output, are tested in three quizzes and a cumulative final. Students in all three of EAPS' concentrations – K-12 administration, higher education, and policy analysis – take the course.[1]

Principle #1: Accessibility

The diverse academic backgrounds and career goals of educational administration students should be kept in mind when selecting instructional materials and assessment tools for quantitative analysis courses (Creighton 2001b, McMillan 2001). For example, EXCEL® is so widely available that many students enter class as proficient users of this spread sheet program (e.g., as a grade or accounting book) and the rest quickly learn the basic functions (e.g., highlighting cells or entering formulas) necessary for the course. While limited in sophistication and flexibility, EXCEL® is a desktop tool to which educational practitioners frequently have access for conducting exploratory data analysis after completing the course. Some textbooks, such as Levine, Berenson & Stephan's (1999) *Statistics for Managers*, provide not only have well-illustrated directions for using EXCEL® but also focus on practical application of statistics to administrative or management problems.[2] A supplementary text, such as Jaeger's (1993) *Statistics: A Spectator Sport*, can help deepen students' understanding of statistical concepts by presenting them with a minimal use of formulas, in a discursive manner, and by focusing on the application of statistics in education. Students are also encouraged to use other supplementary texts such as Dretzke (2002), Gonick & Smith (1993), and Vogt (1999). Whatever texts are used, we recommend that they focus on relatively basic statistical procedures that school leaders might find useful in analyzing student or school data to answer simple questions related to academic performance or organizational management. These books are likely to be accessible by the majority of students so they can develop useful data analysis skills.

Technology and teaching assistants can also be used to help students feel less anxious about the course. In EAPS614, EXCEL® is used to increase students' speed and accuracy in generating descriptive statistics, which they can use in the simple arithmetic for calculating *Z*-scores, confidence intervals, and one-sample tests. Homework assignments include directions and hints so students can be more confident of their empirical results. Not only able to clearly explain statistical concepts and applications, the teaching assistants are experienced practitioners to whom students are able to relate and feel comfortable discussing their frustrations.[3] Smaller class size (not more than 25 students) also allows adequate time for questions concerning key concepts, as well as their demonstration through in-class activities such as those described by Gelman & Glickman (2000). While the various details may vary based on the circumstances, introductory statistics courses can be fairly easily structured to provide students with the resources they need to gain basic quantitative data analysis skills.

Principal #2: Focus on Essential Concepts and Skills

Educational administrators and other practitioners need to be armed with basic knowledge such that they can conduct simple analyses and understand more advanced analyses of statisticians, who often sound to many educational practitioners like they are speaking a foreign language (McNamara 2000). Achieving this goal in one semester requires focusing on fundamental concepts and theories underlying statistical reasoning as well as basic statistical procedures possible with EXCEL®. Focusing on basic distinctions or concepts can highlight for students how the type of data determines the appropriate statistical procedures. For example, students in EAPS614 are expected to understand the difference between the two major families of variables – categorical and scale – and the corresponding descriptive statistics.[4] This distinction is important later in determining whether a χ^2-test for proportions or a *t*-test for means should be used to compare an outcome for two groups. In practice, students should know that the former would be used for analyzing the probability of being classified in a proficiency level (ranging from 1 to 4) on the New York State tests and the latter for analyzing differences in scale scores (ranging up to 840) for the same tests. Even these basic concepts provide students with a foundation to refer to when they encounter more advanced techniques in other courses or in discussions with professional data analysts.

While McNamara (2000) and Creighton (2001b) recommend focusing on exploratory analyses and descriptive statistics, introducing future educational leaders to probability and hypothesis testing can arm them with a healthy skepticism of statistical results. Again, focusing on key issues can help students understand how the concepts and theories relate to each other. For example, the lectures can discuss the links between binomial probabilities, the sampling distribution of proportions, and a *Z*-test to estimate the probability that a certain proportion is obtained by chance given a particular hypothesized true value.[5] For this example, variations in proficiency score rates across schools can be used to frame the discussion of random variation and statistical significance. While one EAPS614 mantra is "Statistics prove nothing," the course is designed to help students understand how statistical significance tests provide evidence upon which to base substantive arguments. Hypothesis tests provide educational leaders with a yardstick for deciding whether a phenomenon, such as changes in average test scores, may reflect random fluctuations or trends potentially related to or influenced by instruction.[6]

Principal #3: Integration of Theory and Practice

Finding relevant examples to illustrate statistical theory and applications is vital for helping student translate what they learn in introductory statistics courses to their role as school leaders (Creighton 2001a, McNamara 2000). Courses intended to build leadership capacity through development

of leadership skills should be structured around a coherent series of analyses of a data set, rather than a disconnected set of statistical procedures. To demonstrate how quantitative data can be used to "tell a story," EAPS614 students are guided through parallel analyses of three data sets – in the lectures, homework assignments, and group project – exploring issues relevant to educational leadership. In addition to information on how to organize and correctly analyze data, quantitative analysis courses for education leaders should also emphasize the need to make substantive interpretations of their results and effectively present them.

When selecting data sets for homework analyses, the topics and variables should be fairly simple and straightforward so students can fairly easily understand the purpose of analyses and develop substantive interpretations of their results. In EAPS614, the homework assignments guide students through analysis of faculty salaries in a hypothetical school district facing union challenges concerning gender inequity in pay. Although the data set was developed for instructional purposes, the variables are similar to those available in a personnel database: salary, gender, highest degree earned, years of service, and whether or not a faculty member also has administrative duties. The assignments use the procedures demonstrated in class to guide students from generation of descriptive statistics to get to know the data to running multiple regression to determine if a salary gap remained after controlling on other variables. Students are strongly encouraged to go beyond simple statistical interpretation to draw substantive implications based on their results. For example, they not only need to correctly interpret descriptive statistics concerning the distribution of salaries in their homework data, but also discuss whether they feel the results indicate that the faculty is "well-paid" or not.

Quantitative data analysis can be like putting together a jigsaw puzzle without a picture on the box cover and where some of the pieces are missing, making assembling the puzzle both very frustrating and also extremely rewarding as the pieces fall into place. To provide experience working with real data, students work in small groups analyzing a data set concerning a topic or problem of interest to them. When the course was first offered, groups were encouraged to use data from their schools or organizations that resulted in some very creative projects varying from evaluations of the physical fitness of third and fourth graders, to students' attendance patterns at an area vocational school, and to use of tutoring services at a local college. To provide a reasonable probability of some statistically significant results, these data sets should consist of at least 100 cases and of around 20 variables. Groups are expected to use approximately 10 of the variables that best address the issue or problem the group is exploring and to employ the appropriate statistical techniques for those variables. Unfortunately, pending Institutional Review Board approval of procedures to deal with human subjects issues, students are currently limited to publicly available data such as New York State school report cards or prepared extracts from national databases.[7]

The final product of the group projects is a brief statistical report, such as a school report card or report to district officials or a school board. One project focused on differences in 4th grade English Language Arts (ELA) examination scores for students in traditional classrooms compared to those in multi-age (split between 4th and 5th graders) classrooms in an area school. Initial analyses indicated that the students in the multi-age classrooms had statistically significant higher average scores than those in traditional classrooms. The group discussed several possible explanations for these findings, such as the influence of parental requests for placement in the multi-age classes or the extra time these teachers invested in coordinating their lessons. While the group projects allow only preliminary examinations of data, they can deepen students' ability to critique statistical studies and lay the foundation for more complex analyses.

In summary, our experience is that a well-structured course, relatively easy to use and widely available software, and time for students to get individualized help allows all students to perform well in the course and gain valuable practical skills. Handouts of lecture notes are provided to students so they can concentrate on the points being covered, ask questions, and fill in comprehension gaps without worrying about tremendous amounts of note taking. Homework assignments directly reinforce key points from the lectures, with a question related to each week's material, and provide an example of a logical series of quantitative analyses. The group project provides an opportunity for students to apply these skills on their own, but with easily accessed advice of the instructor and teaching assistant. The material presented and accompanying exercises provide students with a foundation upon which they can build in other courses or as practitioners. Due to enrollment demand, the number of sections of EAPS614 has been increased with two additional instructors using the same lecture materials and homework assignments.

Uses of Data Analysis by Education Practitioners

One indicator of the utility introductory statistics courses is students' continued use of the skills covered after completing the course. Reflecting the success of the course and demand for data-driven decision making in schools, former EAPS614 students often anecdotally report how they continually use basic statistical skills and techniques to analyze school data as part of their jobs. They are also able to understand and apply statistical procedures discussed in practitioner-oriented journals such as *Practical Assessment, Research & Evaluation* (on-line at http://edresearch.org/pare/home.htm) and participate in discussions with the New York Schools Data Analysis Technical Assistance Group (for more information, see http://www.nysdata.org). Although many researchers would consider their results to be only suggestive due to a lack of rigorous sampling and control variables, these practitioners' analyses provide information for making decisions

concerning local issues related to their student composition, curriculum, and resource allocation. The rest of this article contains examples of the types of analyses undertaken by former EAPS614 students using data from their schools.

Assessing Predictors of Academic Success

Examining relationships among assessment results and other potential predictors of academic success provides insight into current academic programs and student performance (McMillan 2000). The association between various assessments and student characteristics is valuable information that can be used in developing a strategic plan for academic improvement for a particular school. Drawing information from an area middle school's central database, the following are simple analyses of students' academic progress that provided a foundation for faculty discussions related to developing a school improvement plan.

In examining the 8th grade assessments for a recent cohort, a clear association was found between students' scores on the social studies and science examinations (Table 1). The correlation coefficient indicates a relatively weak positive association (r = .32) between students' scores on these two exams, indicating that students who scored higher on one test also tended to score higher on the other. However, the strength of this relationship indicates that students who performed well in social studies may not necessarily experience the same success in science. Thus, efforts to improve students' performance in one subject are unlikely to translate into higher scores on the other assessment.

Table 1

Correlations Assessing Predictors of Students' Academic Success

Variables	Correlation
Social Studies - Science test scores	.32*
1st Quarter – 2nd Quarter grades	.13
1st Quarter – 4th Quarter grades	.87*
Number of absences – 4th Quarter grades	-.51*

* $p < .001$

The academic development of a cohort over the course of the school year can be analyzed by exploring whether students' grades improved, dropped, or remained constant on average over the course of the school year. A comparison of the cohort's first quarter average and the second quarter average in Figure 1 shows a very weak correlation (r = .13) that was not statistically significant. In contrast, the strong positive correlation (r = .87) between the cohort's 1st quarter grades and their performance in the 4th quarter is clearly shown in Figure 2.[8] Despite the fact that students' average grades appear to fluctuate randomly throughout the school year, how they performed at the end of the first quarter is a good indication of how they will perform at the end of the 4th quarter. Administrators can use this data to discuss substantive implications of these results with faculty and staff, such as emphasizing that resources and instructional strategies should be used to ensure the academic success of students through the first ten weeks of school. The results further imply that the current approaches to providing remedial support do not appear to help students who perform poorly at the beginning of the year improve their relative performance by the end of the year.

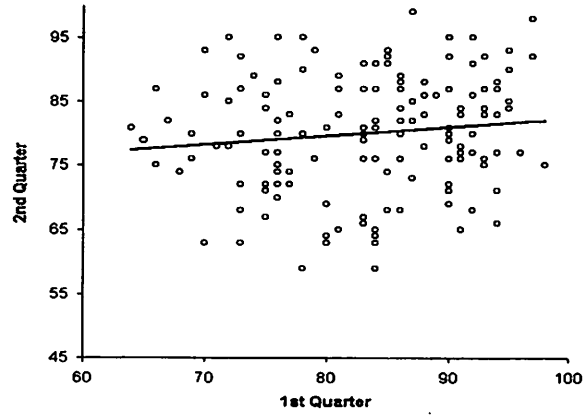

Figure 1: Change in average grades from 1st to 2nd quarter for an 8th grade cohort.

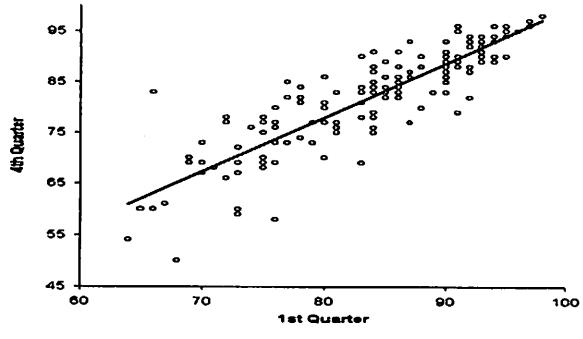

Figure 2. Average change in grades from beginning to end of the school year for an 8th grade cohort.

Anaysis can include other variables that are traditionally viewed as non-academic factors, such as student attendance, that may influence student achievement. For example, what is the relationship between students' daily attendance over the course of the school year and 4th quarter grades? How strongly is missing more school associated with course grades? Again, using data from a recent 8th grade cohort, the trend line in Figure 3 indicates an inverse relationship (r = -.51) between atendance and 4th quarter grades. On average, as students' 4th quarter grades increased, their

days absent from school decreased. Analysis of this type not only indicates the importance of attendance for academic progress, but also has larger policy implications, especially in New York, which has statewide attendance regulations. Based on this analysis, incentive programs can be explored to improve student attendance and deterrents can be implemented to discourage tardiness and absences. While the analysis does not prove a causal link between attendance and academic performance, the results such as these can provide a foundation on which practitioners can base further quantitative and qualitative analyses of their school.

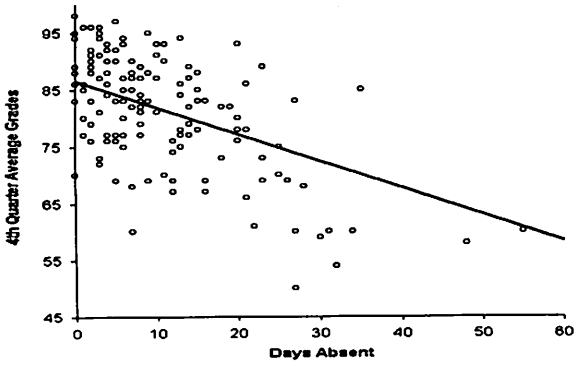

Figure 3. Average end of year grades by number of days absent for an 8th grade cohort.

Evaluation of Remedial Reading Programs

Schools seeking additional resources and financial support often need to include empirical analyses related to their programs in grant proposals. For example, preparation of a Title I grant proposal for an area school district involved evaluating the effectiveness of two academic intervention programs for students struggling in English. One program was a traditional pull-out program in which students worked in small groups on supplementary reading and language arts activities during the school day. The other, newer, program involved similar small groups of students meeting with teachers before or after school for an hour, twice a week, for approximately six to eight weeks. The goal of the statistical analysis was to examine differences in reading achievement growth related to participation in one or both of the programs.

The analyses were complicated because the tests used to assess student learning were different in 3rd grade (CTB McGraw Hill, Terra Nova tests) and in 4th grade (NYS 4th Grade ELA). Using z-scores to measure students' relative standing compared to the mean for their cohort each year, the 3rd grade reading scores on the Terra Nova tests were compared with students' 4th grade ELA score (Russell 2000). For example, students who scored one standard deviation below the mean of their cohort one year would show a .6 standard deviation gain if they scored only .4 standard deviations below the mean the following year. Achievement growth was measured as change in z-scores to indicate whether low achieving students in these programs gained on the average for their cohort. The analysis of the traditional pull-out program combined information from two cohorts for a total of 161 students, while only information from one cohort of 78 students was available for the newer program.

The results from preliminary analyses of average differences in *z*-score growth are shown in Table 2. The mean change in *z*-scores for students in the newer program was .074 compared to -.087 for the rest of the cohort, a difference that was not statistically significant. The mean change in *z*-scores for the remedial reading treatment group was .176 compared to -.077 for other students in the two cohorts, which was marginally statistically significant. Thus, while students in both programs tended to gain on the their cohorts as a whole, only those in the traditional remedial program showed statistically significant relative growth in English test scores. However, the difference between traditional remedial students and the rest of their cohort was not significant in a regression analysis controlling on ethnicity, gender, identified disability, and free lunch eligibility. The lack of significant regression results in this analysis could be due to the small sample sizes, which is often a challenge when attempting to assess the effectiveness of new programs.

Table 2

Evaluation of Remedial English Programs

Type of program	Relative change compared to cohort (*z*-score change) Mean	Std	Sample Size
Before/after school tutoring			
Participated	.074	.891	42
Did not participate	-.087	.967	36
Pull-out remedial tutoring			
Participated	.176*	.834	49
Did not participate	-.077	.848	112

* $p < .05$

Based on the marginally statistically significant differences in average growth, the district considered hiring additional reading instructors to provide intervention services during the school day as part of its Title I grant. The district is also assessing how the before and after school tutorial program might be made effective. Finally, because of the strong relationship between 3rd grade Terra Nova scores and 4th grade ELA scores, the district is using the former to identify students on the border of achieving "proficient" scores for targeted intervention services.

Conclusion

Unfortunately, somewhere between naiveté and intimidation, administrators often view statistics and quantitative analysis as an area that does not concern them or is for academia, not practitioners (Creighton 2001b). Practicing administrators must take advantage of available resources to become proficient in analyzing and interpreting data to improve pedagogical practices and student achievement. Although the expectation has always been for the building level administrator to provide academic leadership for faculty and staff, the current national political and social dynamics regarding education and accountability have emphasized the role of the principal as instructional leader more so than ever over the past decade.

Introductory courses in statistics can not only teach the practicing administrator different methods for manipulating data, but also emphasize how these techniques can be applied in their daily work (Creighton 2001a). This pragmatic approach to teaching statistics goes beyond the typical graduate level course, which is frequently overly saturated with theory and with complicated mathematical procedures that prevent students from understanding the application of those tools in the field. However, unlike Creighton (2001a) and McNamara (2000), we argue that school practitioners need to be armed with more than familiarity with descriptive statistics. A basic understanding of the theory behind inferential statistics arms practitioners with a healthy skepticism of supposedly "hard numbers," which can be especially important when analyzing test score results both at the individual and aggregate levels (McMillan 2000). Once this premise is understood, resourceful educational leaders can take advantage of quantitative data analysis skills to generate information for their building or district regarding student achievement, teacher efficacy, drop-out rates, special-needs learners and many other groups and subgroups that contribute to the daily function of the educational organization. If building level administrators are ultimately accountable for the success of their faculty, staff, and students, it is imperative that they acquire and understand the tools of quantitative analysis in an effort to help them become successful education leaders.

References

Confidence Intervals. (2003). Retrieved September 18, 2003 from the New York State Department of Education's website at http://www.emsc.nysed.gov/irts/school-accountability/confidence-intervals.htm.

Creighton, T. B. (2001a). Data Analysis and the Principalship. *Principal Leadership, 1* (pp. 52-57).

Creighton, T. B. (2001b). Data Analysis in Administrators' Hands: An Oxymoron? *School Administrator, 58* (pp. 6-11).

Dretzke, B. J. (2002). *Statistics with Excel, 2/e.* Englewood Cliffs, NJ: Prentice-Hall.

Fenster, M. J. (1992). Statistics Courses and Adult Learners: Assessing Potential Problems. *Journal of Studies in Technical Careers, 14* (pp. 11-22).

Gelman, A., & Glickman, M. E. (2000). Some Class-Participation Demonstrations for Introductory Probability and Statistics. *Journal of Educational and Behavioral Statistics, 25* (pp. 84-100).

Gonick, L., & Smith, W. (1993). *The Cartoon Guide to Statistics.* New York, NY: Harper Collins.

Jaeger, R. M. (1993). *Statistics: A Spectator Sport.* Second Edition. Thousand Oaks, CA: Sage Publications, Inc.

Levine, D. M., Berenson, M. L, & Stephan, D. (1999). *Statistics for Managers: Using Microsoft® EXCEL.* Englewood Cliffs, NJ: Prentice-Hall.

McMillan, J. H. (2001). *Some Pedagogical Tips for Teaching Statistics.* Paper presented at the Annual Meeting of the American Educational Research Association.

McMillan, J. H. (2000). Fundamental Assessment Principles for Teachers and School Administrators. *Practical Assessment, Research & Evaluation, 7* (8). Retrieved September 18, 2003 from http://edresearch.org/pare/getvn.asp?v=7&n=8.

McNamara, J. F. (2000). Teaching Statistics in Principal Preparation Programs: Part 2. *International Journal of Educational Reform, 9* (pp. 373-384).

New York State Department of Education. (n.d.) *Education Requirements, Exams, Tests, Assessments.* Retrieved September 18, 2003 from http://www.emsc.nysed.gov/deputy/Documents/alternassess.htm.

No Child Left Behind (2001). Retrieved September 18, 2003 from the U.S. Department of Education's website at http://www.nochildleftbehind.gov/.

Russell, M. (2000). Using Expected Growth Size Estimates to Summarize Test Score Change. *Practical Assessment, Research & Evaluation, 7*(6). Retrieved September 18, 2003 from http://edresearch.org/pare/getvn.asp?v=7&n=6.

Vogt, W. P. (1999). *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences,* 2nd Edition. Thousand Oaks, CA: Sage Publications, Inc.

Footnotes

[1]Ph.D. students are required to fulfill their statistics requirement by passing this course or the departmental statistics examination before entering the doctoral research sequence.

[2]Many educational statistics books focus on analyses of experimental data and social science statistics texts often feature analyses of large-scale survey data, both types of data and analyses unlikely to be used by educational leaders interested in improving their schools or school systems.

[3]EAPS has had no difficulty recruiting part-time students as teaching assistants in exchange for one-course tuition reimbursement and credit for an independent study.

[4]More subtle distinctions of sub-types, such as nominal and ordinal categorical variables, are only briefly mentioned so that students are familiar with terms. Statistical procedures where these distinctions are important-such as various types of correlation coefficients-are beyond the scope of this course and the capability of the software used.

[5]EAPS614 is designed as a data analysis course rather than a research methods course. While the data is assumed to be reasonably reliable and representative of the target population, issues of measurement and sampling error are raised as reasons for conducting hypothesis tests but are not discussed in much depth.

[6]To take into account random variation between cohorts when determining whether schools have met the state's annual performance target, New York State Department of Education has published a table of confidence intervals, adjusted to the number of participating students, for various school performance indicators (Confidence Intervals 2003). Such instances of the use of inferential statistics in school accountability systems are examples of why school administrators need to be exposed to more than simple descriptive statistics and exploratory analyses.

[7]Groups select questionnaire items from the National Education Longitudinal Study of 1988 (NELS:88) or Schools and Staffing Survey (SASS), from which subsamples of about 1,000 cases are drawn.

[8]With the principal's permission, these figures are now used as an EAPS614 quiz item asking students to identify which picture matches a particular correlation coefficient.

Kathryn S. Schiller is an Assistant Professor in the Department of Educational Administration and Policy Studies at the University at Albany, State University of New York. In addition to teaching quantitative data analysis, her current research interests include sociology of education, stratification of opportunities to learn, mathematics and science curriculum, and social inequality.

Joseph P. Dragone has worked in public education for 11 years. For the past three years, he has been the principal of Cohoes Middle School, Cohoes, New York. He is a doctoral student in the Department of Educational Administration and Policy Studies at the University at Albany, State University of New York.

Annette Trapini is Assistant Principal at Mayfield Elementary School in Mayfield, NY.

Reliability of Scores and the Researcher

Sean A. Forbes
Margaret E. Ross
Auburn University

As researchers, we want our work and resulting manuscripts to be respected. However, research must be interpretable in order to have value. Thus, the importance of consistency in measuring the variables of interest cannot be overlooked. Without consistency in our measures, we cannot be sure we are making valid interpretations based on the research we've completed. In this article we provide an overview of classical test theory and reliability estimation and provide recommendations for use of reliability estimation procedures that are useful to the researcher. Topics addressed include standard error of the measure, measures of consistency, and inter-rater agreement.

If research is to have any value, it must be interpretable. Recognition of this issue by pioneers of measurement and evaluation led to the development of estimates of the quality of data. Among these, reliability is of note (Spearman, 1910). New students of research quickly learn that reliability reflects a degree of consistency among scores. Teachers hope it becomes apparent to their students that this term has multiple dimensions. Researchers demand it of each other (Spearman, 1910; Pedhauzer & Schmelkin, 1991; Feldt & Brennan, 1993; APA, 1994; Thompson, 1999; Vacha-Haase, Henson, & Caruso, 2002).

Despite repeated, well-articulated calls for attention to reliability, researchers often fail to include reliability information in reporting of data (Feldt & Brennan, 1993; Meier & Davis, 1990; Thompson, 1999; Thompson & Vacha-Haase, 2000). Vacha-Haase, Ness, Nilsson, and Reetz, (1999), for example, reported that among articles published in three professional psychology and counseling journals, nearly 40% of research studies contained no mention of reliability. Of the remaining articles, one-third mentioned reliability scores attained in other studies only. Unfortunately, this is not an isolated finding. Education literature is replete with examples that either misrepresent the nature of reliability or avoid it all together (see e.g., Thompson & Snyder, 1998; Henson, Kogan, and Vacha-Haase, 2001; Yin and Fan, 2000).

But before we lament the downfall of research, we need be reminded of Kierkegaard (1983) who suggested, "Concepts, like individuals, have their histories and are just as incapable of withstanding the ravages of time as are individuals (p. 47)." With this in mind, our intention, then, is to outline foundational notions of reliability in hopes that readers may rediscover the importance of the concept. Reliability estimates reflect the definition of consistency embedded in each of the different forms of reliability. The

[4]More subtle distinctions of sub-types, such as nominal and ordinal categorical variables, are only briefly mentioned so that students are familiar with terms. Statistical procedures where these distinctions are important-such as various types of correlation coefficients-are beyond the scope of this course and the capability of the software used.

[5]EAPS614 is designed as a data analysis course rather than a research methods course. While the data is assumed to be reasonably reliable and representative of the target population, issues of measurement and sampling error are raised as reasons for conducting hypothesis tests but are not discussed in much depth.

[6]To take into account random variation between cohorts when determining whether schools have met the state's annual performance target, New York State Department of Education has published a table of confidence intervals, adjusted to the number of participating students, for various school performance indicators (Confidence Intervals 2003). Such instances of the use of inferential statistics in school accountability systems are examples of why school administrators need to be exposed to more than simple descriptive statistics and exploratory analyses.

[7]Groups select questionnaire items from the National Education Longitudinal Study of 1988 (NELS:88) or Schools and Staffing Survey (SASS), from which subsamples of about 1,000 cases are drawn.

[8]With the principal's permission, these figures are now used as an EAPS614 quiz item asking students to identify which picture matches a particular correlation coefficient.

Kathryn S. Schiller is an Assistant Professor in the Department of Educational Administration and Policy Studies at the University at Albany, State University of New York. In addition to teaching quantitative data analysis, her current research interests include sociology of education, stratification of opportunities to learn, mathematics and science curriculum, and social inequality.

Joseph P. Dragone has worked in public education for 11 years. For the past three years, he has been the principal of Cohoes Middle School, Cohoes, New York. He is a doctoral student in the Department of Educational Administration and Policy Studies at the University at Albany, State University of New York.

Annette Trapini is Assistant Principal at Mayfield Elementary School in Mayfield, NY.

Reliability of Scores and the Researcher

Sean A. Forbes
Margaret E. Ross
Auburn University

As researchers, we want our work and resulting manuscripts to be respected. However, research must be interpretable in order to have value. Thus, the importance of consistency in measuring the variables of interest cannot be overlooked. Without consistency in our measures, we cannot be sure we are making valid interpretations based on the research we've completed. In this article we provide an overview of classical test theory and reliability estimation and provide recommendations for use of reliability estimation procedures that are useful to the researcher. Topics addressed include standard error of the measure, measures of consistency, and inter-rater agreement.

If research is to have any value, it must be interpretable. Recognition of this issue by pioneers of measurement and evaluation led to the development of estimates of the quality of data. Among these, reliability is of note (Spearman, 1910). New students of research quickly learn that reliability reflects a degree of consistency among scores. Teachers hope it becomes apparent to their students that this term has multiple dimensions. Researchers demand it of each other (Spearman, 1910; Pedhauzer & Schmelkin, 1991; Feldt & Brennan, 1993; APA, 1994; Thompson, 1999; Vacha-Haase, Henson, & Caruso, 2002).

Despite repeated, well-articulated calls for attention to reliability, researchers often fail to include reliability information in reporting of data (Feldt & Brennan, 1993; Meier & Davis, 1990; Thompson, 1999; Thompson & Vacha-Haase, 2000). Vacha-Haase, Ness, Nilsson, and Reetz, (1999), for example, reported that among articles published in three professional psychology and counseling journals, nearly 40% of research studies contained no mention of reliability. Of the remaining articles, one-third mentioned reliability scores attained in other studies only. Unfortunately, this is not an isolated finding. Education literature is replete with examples that either misrepresent the nature of reliability or avoid it all together (see e.g., Thompson & Snyder, 1998; Henson, Kogan, and Vacha-Haase, 2001; Yin and Fan, 2000).

But before we lament the downfall of research, we need be reminded of Kierkegaard (1983) who suggested, "Concepts, like individuals, have their histories and are just as incapable of withstanding the ravages of time as are individuals (p. 47)." With this in mind, our intention, then, is to outline foundational notions of reliability in hopes that readers may rediscover the importance of the concept. Reliability estimates reflect the definition of consistency embedded in each of the different forms of reliability. The

degree to which different estimates of reliability estimate potential sources of error is, then, the standard upon which use should be based. It is with this in mind that we address the use of reliability estimates in educational research.

Indices of Reliability

Generally, in terms of quantitative estimates, reliability reflects the assumptions of classical test theory. Literature concerning this area is voluminous, and details of this issue are beyond the scope of this article. But, readers are encouraged to review Crocker and Algina (1986), Lord and Novick (1968), Nunnally (1978), and Stanley (1971), among others, for thorough discussions of these foundational issues.

Classical test theory gives rise to two forms of variance partitioning estimates of reliability: (1) estimates developed with use of the standard error of measurement (S.E.M.) and (2) estimates that result in reliability coefficients. Each form has utility but both are not often used in tandem. Ideally, choice of reliability estimates should not be either/or.

Standard Error of Measurement

Standard error of measurement refers to the standard deviation of error in a data set. It serves as an estimate of reliability by providing a confidence interval for scores. As a result, the most significant quality of the S.E.M. as an estimate of reliability is in its interpretability as results are given in score oriented units. In this regard, researchers can determine the amount of flexibility in an individual's score. Another, although more complex, advantage to using S.E.M. based estimates of reliability is the ability to partition various error sources. From this information researchers are provided practical information on how instrumentation can be improved.

The S.E.M. is calculated by the following formula: multiplying the standard deviation of a set of scores by the square root of 1 minus the reliability coefficient:

$$s_x \sqrt{1-p}$$

where s_x is the standard deviation of a set of scores and p is a reliability coefficient estimate.

Consider an example where the standard deviation of IQ scores was 5.2 and one form of reliability estimate of the scores from the instrument was reported as .89. Using the formula above the
S.E.M. is

$$5.2 \sqrt{1-.89}$$
$$5.2 \sqrt{.11}$$
$$1.72$$

Researchers can then multiply the S.E.M. by a Z-score for a specific alpha level:

(S.E.M.) (Z-score of 95% interval)

This results in a range of scores in which researchers can confidently conclude the observed score exists:

$$(1.72)(2)$$
$$+/- 3.44$$

For example, let us say that a student received a score of 102 on an intelligence battery. Using the quotient developed by multiplying S.E.M. with a Z-score at the 95% interval, we can conclude with 95% confidence that the student's true score was between 98.56 and 105.44.

Reliability Coefficients

According to current interpretations of classical test theory, a reliability coefficient is represented by the formula:

$$s_t^2 / s_e^2$$

where s_t^2 is the variance of true scores and s_e^2 is the error variance of scores. Following this formula, the degree to which there is less error in measurement is related to higher levels of reliability. Resultant coefficients range from 0 to 1.0, a higher value indicating a greater degree of reliability than lower values. It is this characteristic that makes coefficient most useful. Such a standardized metric allows estimates of reliability to be compared across data sets, with no concern for units of measurement. This, however, is where similarity among reliability coefficient estimates ends. As coefficient estimates assess different sources of potential error, awareness of the inability to use methods interchangeably is necessary.

Within educational research, the most common forms of coefficient estimates of reliability include: (1) alternate form, (2) test-retest, (3) internal consistency, and (4) inter-rater. Each will be addressed in turn, though, among these, the latter two are more evident within educational research literature than the former. It is important to note that, according to classical test theory, means and error variances of test items or parts must be equal to satisfy the conditions of parallel forms, which is assumed when calculating alternate form and test-retest reliability estimates.

Test-rest

Let's reconsider our student who completed an intelligence battery. Say that four weeks after the first assessment the individual completed the same activities again and responses were scored. It's extremely likely that the value of the first measurement will differ from that of time two, at least marginally. This difference, however, is not a difference in true score, but a difference in observed score attributed to sources of error introduced by the different times of measurement. If every condition of time 1 were replicated during time 2, there would be no difference in observed scores between assessments.

Such a situation is known as a test-retest estimation of reliability. The most significant benefit of this form of estimation is the information it provides on a single assessment of interest. If researchers can be reasonably sure that conditions in time 1 and time 2 are similar, low reliability estimates point to a test that is not an accurate

operationalization of a construct.

Two caveats, however, are warranted in using a test-retest estimate. First, the trait or variable of interest cannot be assumed to change from time 1 to time 2. Secondly, memories of those being assessed during time 1 cannot influence the performance on the second assessment. Whereas these cautions are of little concern to researchers considering issues such as physical performance, IQ and the like, it is a major limitation when assessing reliability of scores of content knowledge or cognitive and/or emotional states.

Probably the largest limitation of test-rest estimates is an issue of practicality. Assessment must be completed on multiple occasions. And, given the constant struggle for resources in education, most teachers, principals, and superintendents are rarely amenable to two bites of the proverbial student apple.

Alternate Form[1]

Similar to test-retest estimates of reliability, alternate form estimates require respondents to complete two assessments. With alternate form estimates, however, respondents complete two forms of an assessment, typically during one testing session...not at different times as with test-retest estimates. Alternate forms can also be employed over time, however, users are advised to account for the added sources of error.

Using our previous example, an alternate form estimation of reliability would require our respondent to complete two intelligence batteries at one sitting. Again, error in measurement would account for differences between scores from test 1 and test 2. But, with this form of estimation, error would largely be a result of using alternate forms, not time.

Alternate form estimates of reliability are useful to researchers in that they do not necessarily require multiple assessment times. Moreover, estimates provide researchers with an index to measure the consistency with which two instruments assessed a construct. Since time in not a factor if completed in one setting low reliability estimates point to a lack of consistency between measures, something that does not bode well for drawing conclusions based on such data. Another benefit of these estimates is the ability to assess a construct that is not stable over time...states of emotionality, immediate cognitive responses, etc. And, since different forms are used one instrument tends not to inform completion of another form.

Selecting alternate form of reliability estimation depends on several factors. First, alternate forms must be generated or available. This is not to say that alternate forms are different forms. Instead, alternate forms are to be statistically equivalent under the assumption of classical test theory. Also, researchers must have confidence in one of the forms of measurement. Without it, reliability coefficients simply suggest assessments are measuring some trait with consistency—not necessarily the trait you believe is being measured. Finally, researchers must be able to counter balance administration of instruments so as to avoid fatigue or ordering effects.

Taken together the effort involved in completing an alternate form estimation of reliability makes it less than attractive to many researchers, especially those in education. As with test-retest estimates, having students or other respondents complete two estimates of a construct is a strain on available resources.

Internal Consistency

Internal consistency estimates measure the degree to which parts of an assessment scale uniformly addresses the same construct. For example, if our scale measures verbal ability, all parts of the scale should address verbal ability. In this section, issues related to internal consistency estimates of reliability are discussed.

Because of the logistic problems with arranging for a substantial number of people to respond to an instrument two times and difficulties with developing a parallel form of an instrument, test-retest and alternate-forms approaches to estimating reliability are not often used by researchers. Thus, researchers rely mainly on internal consistency approaches to estimating reliability. One major advantage to internal consistency approaches to estimating reliability is that the instrument has to be administered only once. Considering that traits may change across time, reliability estimates at one point in time can even be regarded as preferable to those that require a lapse in time (Feldt & Brennan, 1993).

In order to assess internal consistency, an instrument must be divided into two or more parts that represent different test forms so that the parts can be compared as in alternate-forms approaches. The Spearman-Brown formula is the oldest reliability estimation procedure when two parts are used (Feldt & Brennan, 1993). The calculation for the Spearman-Brown is

$$(2)\ r\,\chi_1\chi_2 / 1 + r\,\chi_1\chi_2$$

where χ_1 and χ_2 represent the two part scores and $r\,\chi_1\chi_2$ is the correlation between the two scores. To illustrate, suppose a researcher developed an instrument to measure attitudes toward comprehensive examinations with 10 items measured on a Likert-type scale. The researcher decided to assess reliability of scores from the instrument through using the Spearman-Brown. First, he or she would first decide how to divide the instrument into two parts. We will suppose the researcher decided to compare odd and even items instead of first five and last five items. Second, the researcher calculates a score for odd items (e.g. a summed score) and a score for even items. Third, the scores for the two parts are correlated. Fourth, the correlations are used in the formula for the Spearman-Brown.

However, the Spearman-Brown assumes that the two parts are parallel in the classical sense, thus consideration must be given to item writing and how the instrument or the instrument's scales are divided so that the parallel-forms requirements are met. A particularly daunting problem is writing items so that the error variances are equal and so

that the parts yield equal means and variances. Due to these difficulties, the Spearman-Brown formula has been modified or other reliability estimation formulas developed to accommodate less stringent requirements.

Tau-Equivalency. Tau-equivalent parts maintain equal means across parts but item error variances and part variances are not equal. Essentially tau-equivalent parts, in addition to displaying differences in error variances and part variances, display part mean differences within an individual score constant. See Felt and Brennan (1993) for a full discussion of Tau-Equivalency. Thus, the part means are equal if some specific value (e.g. 2 points) is added to each individual score. The most common internal consistency reliability estimation used with essentially tau-equivalent parts is Cronbach's alpha (1951). As can be seen in the formula below, when calculating Cronbach's alpha, each individual item error variance is considered:

$$(n / n - 1)\,(s^2_x - \Sigma^2_{xi} / s^2_x)$$

where n is the number of items, s^2_x is the test variance, and s^2_{xi} equals the item variances. Given that the formula for Cronbach's alpha entails calculating item level error variances (s^2_{xi}), hand calculations are quite unwieldy. Although many statistical packages (e.g. the Statistical Package for the Social Sciences, SPSS for Windows Release 11.0.1; 1999-2001) can be used to calculate Cronbach's alpha, it may be necessary to hand calculate the reliability estimation at times. This situation may occur when sample specific reliability coefficients are not reported in journal articles but are needed for research purposes. For example, reliability generalization, a recent meta-analytic procedure for assessing reliability and sources of error (Vacha-Hasse, 1998) relies on obtaining or figuring sample reliabilities for each study. For dichotomously scored items, an alternative reliability estimation, which is feasible to hand calculate from reported means and standard deviations, is the Kuder-Richardson 21, or KR-21 (Kuder & Richardson, 1937). The KR-21 is calculated using the following formula:

$$\left[\frac{n}{n-1}\right]\left[1-\frac{\bar{\chi}_x(n-\bar{\chi}_x)}{(n)s^2_x}\right]$$

In the formula above, χ_x is the instrument/scale score mean, n is the number of items, and s^2_x is the scale's variance.

Congeneric Parts. Congeneric parts are linearly related but relax assumptions associated with parallelism even more than does essentially tau-equivalence (see Feldt & Brennan, 1993, for a fuller explanation of congeneric forms). Kristof (1974) points out that Cronbach's alpha is a lower-bound estimate of reliability stating that "we can do better" (p. 492). He presented a method of reliability estimation using any three content-homogeneous parts that can be hand calculated from the part covariances, which may be easily obtained using any statistical package. Simply divide the scale into three parts, using a statistical package, find the variance for the full scale and covariances for all combinations of parts, and estimate the reliability using the Kristof reliability estimation below:

$$\frac{(Cov_{12}\,Cov_{13} \quad + \quad Cov_{12}Cov_{23} \quad + Cov_{13}\,Cov_{23})^2}{(Cov_{12})(Cov_{13})(Cov_{23})(s^2X)}$$

where Cov is the covariance, the subscripts represent the parts that are covaried, and s^2X is the full scale variance. Since Cronbach's alpha is considered to be a lower-bound estimate and may be a depressed estimate when assumptions related to essentially tau-equivalence are not met, the Kristof reliability estimation could yield a higher value. Thus, it may be to the researcher's advantage to report the Kristof reliability estimation as well as Cronbach's alpha if Cronbach's alpha yields a somewhat low coefficient. However, as Feldt and Brennan (1993) indicate, large differences in the two coefficients should not be expected.

Inter-rater Consistency

A seemingly simple notion, estimating inter-rater consistency has provided a long-standing confusion among consumers of research (Frick & Semmel, 1978; McGaw, Wardrop, and Bunda, 1972; Medley & Mitzel, 1958, 1963; Tinsley & Weiss, 1975). The problem revolves around two statistically related terms...inter-rater reliability and inter-rater agreement.

Consider an example. Two raters view 20, videotaped demonstrations of children's behavior and assign ratings on a ten point Likert-type scale. One rater assigns a '2' to each demonstration; the other rater assigns a '10' to each. The 20 pairs of ratings are perfectly correlated. Reliability between raters is had in its purest form. If this was generated from a data set of test scores, researchers would be ecstatic; but with the use of raters, it is not enough to merely be consistently different. Instead, inter-rater agreement must be considered if there is to be a meaningful reliability with the use of raters. After all, how much confidence can a researcher have in ratings if raters cannot agree on what they are observing, no matter how reliably they do it? High reliability does not guarantee absolute agreement among raters and low reliability does not guarantee that raters absolutely disagree. The result of this twist on reliability is that many researchers settle for estimating inter-rater agreement or inter-rater reliability when prudence would suggest a mutual consideration of terms. Consideration of each follows.

Inter-rater Agreement

Towstopiat (1984) suggests that calculating the percentage of agreement is the most widely used measure of inter-rater agreement due to its ease of calculation and

interpretation. An added advantage of this method is the ability to estimate inter-rater agreement for continuous and nominal data sets. The formula is represented by:

# of Agreements/ Total # of Observations

This approach, however, is hampered by several limitations, including a failure to account for random agreement, the potential of high-frequency behaviors to provide a biased estimate of agreement compared to low-frequency behaviors, and lack of an acceptable lower bound estimate (Cohen, 1960; Freck & Semmel, 1978; Hartmann, 1977; Tinsley & Weiss, 1975; Towstopiat, 1984). Taken together, calculating inter-rater agreement via percentage of agreement is not recommended (Towstopiat, 1984).

Other measures for estimating inter-rater agreement for continuous data sets have been suggested, including multiple chi-square based measures, pair-wise correlations between raters, and Kendall's coefficient of concordance yet Tinsley and Weiss (1975) suggest each have proven inappropriate for estimating inter-rater agreement. Instead, the authors suggest the use of the chi-square statistic proposed by Lawlis and Lu (1972). See Tinsley and Weiss (1975) and Lawlis and Lu (1972) for further a detailed discussion of this issue. More recently, Berry and Mielke (1988) and Conger (1985) suggest the use a generalization of an estimate once reserved for nominal data, specifically Cohen's (1960) Kappa statistic.

Among methods for estimating inter-rater agreement for nominal data, none are more prevalent than the Kappa statistic and respective generalizations. Responding to the inability of researchers to account for chance agreement when using percentage of agreement as an estimate of inter-rater agreement, Cohen (1960) developed the Kappa statistic. It is represented by the formula:

$$Po - Pc/1-Pc$$

where, Po is proportion of agreement between raters and Pc is proportion of agreement expected by chance. A weighted version of Kappa is available in Cohen (1968) for researchers who do not assume any difference in agreement to be of equal interpretive value. Limited by situations involving only two raters, generalizations of Kappa have been developed...i.e. Fleiss's (1971) formula for multiple raters, Kvalseth's (1991) formula for asymmetric distributions, Berry and Mielke's (1988) formula for interval measurement and multiple raters.

Inter-rater reliability

For estimating inter-rater reliability with continuous data, the intra-class correlation is the index of choice (Bartko, 1966, Ebel, 1951; Fleiss, 1975; Shrout & Fleiss, 1979; Tinsley & Weiss, 1975). Variance values generated through analysis of variance procedures allow for estimate of the intra-class coefficient represented by the formula (Bartko, 1966):

$$s_r^2 / s_r^2 + s_e^2$$

where s_r^2 is rater variance, and s_e^2 is rater error variance. The analysis of variance equivalence is represented by the following formula (Bartko, 1966):

$$MSP-MSE / [MSP + (k+1)MSE]$$

Resulting values range from –1.0 to 1.0; however, negative correlations are rarely obtained and indicate rater X ratee interaction (Tinsley & Weiss, 1975).

Despite wide-spread use of intra-class correlations as estimates of reliability, researchers are often unaware of the differences among forms of ICC, or even that there is more than one form. Shrout & Fleiss (1979) identify three questions that researchers must consider when calculating an ICC: (1) Does the arrangement between raters and targets require a one way or two way analysis of variance, (2) Are differences among means of rater's ratings of interest in estimating reliability, and (3) Is the desired unit of analysis individual ratings or means of ratings. For those responsible for selecting the form of ICC, the answers to these questions are to found in the arrangement of the study. That is, as different models exist of how raters will assess targets, different mathematical formulas must be used in estimating the general formula of the ICC (Frick & Semmel, 1978; Shrout & Fleiss, 1979; Tinsley & Weiss, 1975). See Bartko (1966; 1976), Shrout and Fleiss (1979), and Tinsley and Weiss (1975) for detailed discussions of necessary steps and further considerations.

For estimating inter-rater reliability for nominal scales, researchers encounter a unique situation. Tinsley and Weiss (1975) define inter-rater reliability as the "degree to which the ratings of different judges are proportional when expressed as deviations from their means (p. 359)." Problem is, many times researchers earnestly attempt to estimate inter-rater reliability estimates for ratings on nominal data—demonstration or non-demonstration of misbehavior, classification into one of three possible types of readers, etc. For these ratings, deviations from a mean are meaningless. Raters either agree or do not; degree is not assessed as it is with continuous data. For nominal data, then, inter-rater reliability is determined solely by inter-rater agreement.

Guidelines for Selection

A casual review of the previous sections could easily lead one to believe that selecting appropriate estimates of reliability for a research project is a daunting experience. And, if the reporting practices found in current literature are index, it is. Yet, a closer reading reveals only a few elements of instrumentation are necessary to complete an apparent Herculean task.

First, researchers must be mindful of the kind of measures selected for use: Is data collected through self-report or external raters? Next, form of data needs to be considered: Is data continuous, dichotomous, or nominal? Once these questions have been answered selection of a general form of reliability estimation is a foregone conclusion.

For self-report data internal consistency estimates should be the first choice for researchers. Test-retest and alternate-form estimation methods could be used to lend

credence to internal consistency estimates, but the potential drain on resources leaves these options unattractive to researchers for most cases. The caveat to this, of course, is that internal consistency estimates cannot be appropriately used for nominal data. Here, test-retest and alternate form methods of reliability estimation would be appropriate. However, estimating reliability for nominal data is unnecessary unless a researcher believes study participants are deceptive. Even then, an index of social desirability would be a more efficient estimator. It is for these reasons that Table 1 lists only internal consistency indexes for dealing with self-report data. Choice of index is considered in light of the underlying notion of parallelism and form of data.

Table 2 outlines the available estimation indices for inter-rater consistency. Consideration of what type of consistency is needed and the form of data leads researchers to use either the Kappa statistic or the ICC. This, however, is only a starting point for researchers. Numerous considerations are left, and only a researcher's familiarity with a study should direct these (see discussion above).

The Researcher's Responsibility

Once an educated decision has been made regarding the necessary form(s) of reliability estimation, researchers have several responsibilities. At a bare minimum, researchers have an obligation to appropriately report reliability estimates. Yet, the greatest potential benefit of reliability estimation is the information it provides active researchers in adjusting efforts before research conclusions have been made.

Reporting Reliability Estimates

In all cases, whether or not the researcher developed the instrument, refined an existing instrument, or used an existing instrument without changes, a reliability coefficient should be calculated for each scale and reported in any document describing the results of research using the scale (Thompson, 1994). In reporting the type of reliability coefficient should be given along with the value of the coefficient. Additionally, the reliability coefficient should be interpreted briefly. For example, a researcher might say:

> Internal consistency was assessed using Cronbach's alpha, yielding a coefficient of .82. Thus, interpretations based on the deep study-strategy scale scores can be made with acceptable to good reliability.

It is also important to be careful with language when writing about reliability. That is, instruments are not reliable

Table 1

Reliability Estimation Indexes for Self-Report Data: Conditions X Data Form

	Classic Halves	Congeneric	Tau-Equivalent
Dichotomous	N/A	N/A	KR-21
Continuous	Cronbach's Alpha Spearman-Brown Kristof	Cronbach's Alpha	Cronbach's Alpha

Table 2

*Inter-rater Consistency Indexes Appropriate for Consistency Form X Data Form**

	Inter-rater Agreement	Inter-rater Reliability
Nominal	Kappa	N/A
Continuous	Kappa	ICC

*Information contained in this table is designed for initial consideration of inter-rater consistency estimation only. Several conditions must be considered before the final form of estimation can be accurately identified.

(Thompson & Vacha-Haase, 2000) and should not be described as such. Instead, reliability refers to the scores from the instrument for a specific data collection. This is no semantic quibble. An accurate representation of what reliability is and is not is essential to an appropriate interpretation of research findings. After all, if researchers appreciated this issue it would be rare indeed for a published research study to contain claims of reliability for a current study based on data collected years before by another researcher.

Using Reliability Estimates to Improve Instrumentation

Data from social science research, including education, is collected through knowledge, attitude, and performance-based instruments. Often, these come in the form of questionnaires. Other times researchers employ raters to observe the demonstration of a construct of interest. Yet, regardless of instrumentation, reliability estimates provide researchers with information that can lead to an immediate improvement in measurement of a construct.

Questionnaires

Once data is collected, the researcher should calculate the reliability for the sample used. The researcher can simply enter item level data into a statistical package such as SPSS (the Statistical Package for the Social Sciences, SPSS for Windows Release 11.0.1; 1999-2001) and obtain the reliability coefficient. Additionally, item level data is available and the researcher can obtain information about the reliability if individual items are deleted. For example, if item level data indicates that item number five on a scale lowers the overall reliability coefficient for the scale, that item should be reviewed and consideration given to omitting the item.

The researcher should make every effort to refine any instrument he or she developed or revised such that acceptable reliability is obtained. When consideration is given to omitting items, the researcher must consider both statistical data and the theoretical basis for including the item before a decision is made. Once the instrument is refined, the researcher should base further analyses on the final form of the instrument and again calculate the reliability coefficient for reporting purposes. Very generally speaking, for any knowledge-based instrument a coefficient of .90 or higher and for attitude scales, a coefficient of .70 or higher are considered acceptable. However, it is cautioned that no set standards are appropriate for judging the adequacy of reliability coefficients. In some cases, lower reliability coefficients are acceptable but in other cases, higher reliability coefficients are needed. Most textbooks (e.g. Thorndike, 1997) provide a discussion of the appraisal of reliability coefficients. Suffice it to say, the higher the reliability, the more we can be sure that our research conclusions are appropriate.

Raters

Refining questionnaires is a relatively straightforward task, once you know the steps. But, refining a rater is a bit more difficult than refining a 20-item, self-report instrument. This leaves researchers in a tight spot when it comes to discovering that ratings have a low degree of agreement or reliability after the data has been collected. The answer, unfortunately, lies in researchers' abilities to avoid the situation altogether. Agreement among raters regarding the constructs under consideration must be had prior to measurement; without it, reliability and agreement are rarely evidenced when asking raters what has been observed. Here is where researchers can see the benefit of pilot testing their raters. Because, unlike quantitative, self-report data, data collected through raters cannot be partitioned after the fact.

Reliability Beyond the Researcher

Though our immediate interest is the use of reliability estimates among research practitioners, we are motivated to end with a call to the broader community of educators: Reliability is not for researchers alone. A number of groups have significant reason to familiarize themselves with the ins and outs of reliability estimation, including journal editors, consumers of educational research, and educators, in general. For journal editors and reviewers, an understanding of these issues is paramount if legitimate review can be had. Research, no matter how well intended, is ultimately not interpretable void an explicit statement of how reliability was assessed. It is reviewers' and editors' responsibility to ensure that these issues are addressed. For consumers of educational research outside researchers themselves, teachers and parents for example, an understanding of reliability provides a note of caution in interpreting results. Two groups of researchers may conclude very similar results. That is not to say that they have equally compelling evidence. And, among issues of evidence, reliability is as big as it gets.

Conclusion

Inasmuch as reliability of scores is not a one-dimensional construct, varying conceptions of the term are available for use. This paper has focused on estimates of reliability estimated through variance partitioning, namely that which are based on a standard error of measurement or result in a reliability coefficient. And, as reliability coefficients are interpretable without an appreciation of a metric of measurement, as it is not with S.E.M. based estimates, these forms of reliability estimation are most commonly used in social science research.

Currently, four conceptions of reliability estimation result in reliability coefficients—alternate form, test-retest, internal consistency, and inter-rater. Selection of estimation type to be used is contingent upon the type of data collected and the form of instrumentation selected. Researchers, however, are cautioned not to rely on one form over another. Instead,

multiple measures must be had, if at all possible, to reflect a most accurate conception of reliability.

References

American Psychological Association (1994). *Publication manual for the American Psychologist Association* (4th ed.). Washington, DC: Author.

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports, 19*, 3-11.

Berry, K. J., & Mielke, P. W. (1988). A generalization of Cohen's kappa agreement measure to interval measurement and multiple raters. *Educational and Psychological Measurement, 48*, 921-933.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Conger, A. J. (1985). Kappa reliabilities for continuous behaviors and events. *Educational and Psychological Measurement, 45*, 861-868.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378-382.

Hartmann, D. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavioral Analysis, 10*, 103-116.

Henson, R. K., Kogan, L. R., & Vacha-Hasse, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*, 404-420.

Kierkegaard, S. (1983). The Concept of Irony. New York: Octagon Books.

Kvalseth, T. O. (1991). A coefficient of agreement for nominal scales: An asymmetric version of kappa. *Educational and Psychological Measurement, 51*, 95-101.

Lawlis, G. F., & Lu, E. (1972). Judgment of counseling process: Reliability, agreement, and error. *Psychological Bulletin, 78*, 17-20.

McGaw, B., Wardrop, J. L., & Bunda, M. A. (1972). Classroom observation schemes: Where are the errors? *American Educational Research Journal, 9*(11), 13-27,

Medley, D. M., & Mitzel, H. E. (1958). Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behavior. *Journal of Experimental Education, 27*, 23-35.

Medley, D. M., & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand-McNally.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology, 37*, 113-115.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 2*, 420-428.

Thompson, B. (1999, February). Understanding coefficient alpha, really. Paper presented at the annual meeting of the Educational Research Exchange, College Station, TX.

Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent JCD research articles. *Journal of Counseling and Development, 76*, 436-441.

Tinsley, H. E. A., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgements. *Journal of Counseling Psychology, 22*(4), 358-374.

Towstopiat, O. (1984). A review of reliability procedures for measuring observer agreement. *Contemporary Educational Psychology, 9*, 333-352.

Vacha-Hasse, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*(4), 562-569.

Vacha-Hasse, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*, 335-341.

Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educational and Psychological Measurement, 60*, 201-223.

Footnote

[1]Within reliability estimation literature alternate form is referred to as parallel forms. We use the term alternate form, instead. This helps to avoid confusion when reading about notions of parallelism with internal consistency estimation methods.

Sean Forbes is an Assistant Professor in the Department of Educational Psychology/Assessment at Auburn University. His research interests include developmentally appropriate assessment, adolescent development, and teacher education.

Margaret Ross is an Associate Professor in the College of Education at Auburn University. Her research interests include the role assessment plays in influencing student learning strategies, assessment issues and practice, and educational program evaluation.

Author Note

Direct correspondence to Sean A. Forbes, 4036 Haley Center, Auburn University, (334) 844-3083, forbesa@auburn.edu.

Learning Behaviors in a College Setting: A Psychometric Examination of Scores on the Learning and Study Strategies Inventory (LASSI) in a Sample of College Students

Barbara A. Schaefer
The Pennsylvania State University

Frank C. Worrell
University of California, Berkeley

The Learning and Study Strategies Inventory (LASSI; Weinstein, Schulte, & Palmer, 1987; Weinstein, Palmer, & Shulte, 2002) is a normative measure used for diagnosing college students' learning problems. In this study, we examined the internal consistency and predictive validity of LASSI first edition scores in 255 college students taking an educational psychology course. We also examined the structural validity of LASSI item and subscale scores in the sample, as well as differences in LASSI scores between students with high and low final grades in the course. Results indicated that internal consistency estimates were in the moderate to high range, LASSI scores predicted student achievement, and students with high grades obtained significantly higher scores on the majority of LASSI subscales than their counterparts with lower grades. The study also confirmed the three-factor structure for LASSI subscale scores and identified a matching three-factor structure for item scores, a finding that has not been reported in the literature previously.

As colleges and universities continue to attract individuals with varying levels of skills, as well as non-traditional and returning adult learners, these institutions are finding that an increasing number of students require educational interventions to be successful. A survey by the Cooperative Institutional Research Program indicated that in 1998, 9% of all entering first year college students reported having a disability ("1 in 11 Freshmen", 2000). Moreover, this survey revealed that a growing number of university matriculants who identify themselves as having a disability indicate they have a learning disability (41% in 1998, as compared to only 15% in 1988). Given these data, a mounting concern at the university level is how to identify what students' learning difficulties are, both for those identified as learning disabled and for students who are not found to meet the criteria for specific learning disabilities or other exceptionalities, but who are nonetheless experiencing academic difficulties.

In the K-12 system, often a standard part of the assessment of students with learning difficulties is an evaluation of their learning behaviors. Learning behaviors are behaviors that are related to academic success, including attention, persistence, and use of strategies, among others. Students who have appropriate learning behaviors typically have higher achievement than students who do not have appropriate learning behaviors (McDermott, 1999). Learning behaviors have been identified in the research literature from as early as the 1970s (e.g., McKinney, Mason, Perkerson, & Clifford, 1975), but systematic measurement of these behaviors did not occur for several years.

Since learning behaviors have typically been studied in students who were not performing well in school, early measures of learning behaviors were developed for students in elementary and secondary schools (e.g., Reynolds, 1979; Stott, Green, & Francis, 1983; Stott, McDermott, Green, & Francis, 1988). Indeed, one can argue that students who are performing well enough in secondary school to go on to college are not students who are in need of interventions to enhance their learning behaviors. However, not all students who attend college are successful. As the breadth of students attending college has increased over time, so has the number of unsuccessful students, as well as the number of students who were in need of interventions related to learning (Mealey, 1988). The increasing enrollment in college of students who have been diagnosed with exceptionalities related to learning ("1 in 11 Freshmen", 2000) has also resulted in an increased need for colleges to address behaviors related to effective learning. In fact, some universities now have regular courses for freshmen on learning to learn (e.g., Pintrich, McKeachie, & Lin, 1987; Schumacher, Sayler, & Bembry, 1995).

The Learning and Study Strategies Inventory (LASSI)

Weinstein (1982) developed and later published the first version of the LASSI (Weinstein, Schulte, & Palmer, 1987) to address the growing need for a scale that examined behaviors and attitudes related to academic success at the college level. The 77-item first edition of the instrument is a self-report measure of "college students' use of learning and study strategies" (Weinstein et al., p. 2) designed to provide a standardized measurement of students' approaches to their academic work. The LASSI purports to measure 10 factors related to student academic success. Five of these are affective in orientation (Attitude, Motivation, Time Management, Anxiety, & Concentration) and five focus on the use of study strategies (Information Processing, Study Aids, Self-Testing, Test Strategies, & Selecting Main Ideas). In the User Manual, Weinstein (1987) indicated that the LASSI can be used both to diagnose and remediate students' weaknesses in the areas measured, and she recommended its use in a pre-intervention/post-intervention fashion to assess students' developing skills. Subsequent to the current study, the 80-item second edition of the LASSI was published and included item changes but retained the same 10 scales (Weinstein & Palmer, 2002).

Since its publication, many colleges have adopted the LASSI (e.g., Olaussen & Bråten, 1998; Robertson, 1994), in part because of the claim by the test manual that the instrument has national norms. In fact, the LASSI first edition norm sample consisted of incoming first year students at a

single institution – clearly not a nationally representative normative sample. The LASSI second edition (Weinstein, Palmer, & Shulte, 2002) includes improved national norms; however, the representativeness of this norm group remains unclear. Mealey (1988) reviewed the instrument and commended the authors on their attempts to choose items that did not result in socially desirable responses. She also noted that the evidence of the reliability of the scores—both internal consistency and test-retest—was "strong" (p. 384). However, Mealey cautioned users of the LASSI that there was little evidence of validity for the instrument's scores and the norm sample did not consist of students with learning problems. Mealey concluded that the LASSI should be viewed as experimental and recommended additional research to establish the validity and generalizability of the scores.

Group differences on LASSI scores. Since Mealey's observations, several studies have been conducted on the LASSI. McKeachie, Pintrich and Lin (1985) taught a one-semester introductory psychology course covering topics related to learning, such as memory, problem solving, and time management. Course participants were in their first academic year. Using the first version of the LASSI (Weinstein, 1982), the researchers reported a positive correlation between the total LASSI posttest score and later GPA, and greater increases in LASSI total scores after the class for the enrolled students than for the students in two comparison groups.

Nist, Mealey, Simpson, and Kroc (1990) examined the first edition of the LASSI (Weinstein, 1987; Weinstein et al., 1987) using it as both pre- and posttest for two groups of students in a course on study strategies. One group was comprised of regularly admitted students who volunteered for the class, and the other group consisted of college students who were required to take the class. Nist et al. reported that the regularly admitted students obtained significantly higher posttest means on all of the LASSI subscales whereas the developmental group obtained higher scores on 8 of the 10 subscales. However, although LASSI subscales accounted for 20% of the variance in the GPA of regularly admitted students, for the developmental college students "no scale or combination of scales was predictive of course grades" (Nist et al., p. 48) in regular courses or in the study strategies course.

Another investigation by Deming, Valeri-Gold, and Idleman (1994) examined first edition LASSI scores in a sample of developmental college students identified as needing study skills and other intervention strategies. Since the LASSI manual indicates that students who score below the 75th percentile will need instruction in learning and study skills, Deming et al. hypothesized that these pre-identified students should score below the 75th percentile on the LASSI. They found that students in their sample scored better than expected on three of the subtests, and no differently from the norms on four subtests. More recently, Landrum and Chastain (1998) examined differences on first edition LASSI scores after a tutoring intervention using class tests and LASSI scores as the dependent variables. Students who received tutoring performed significantly better on later tests relative to the students who did not receive tutoring, but LASSI pre-tutoring and post-tutoring scores did not differ. In another study, Schumacker et al. (1995) also reported mixed findings using LASSI gain scores.

Internationally, Albaili (1997) examined LASSI score differences in groups determined by academic achievement in a university in the United Arab Emirates. Participants consisted of undergraduates who were divided into three groups on the basis of their grade point average (GPA). Using an Arabic translation of the LASSI, Albaili found that the low-achievement group obtained significantly lower scores than the medium- or high-achievement groups on the 10 subscales. A stepwise discriminant function analysis indicated that three subscales best classified the students into low and high achieving groups: Motivation, Information Processing, and Selecting Main Ideas.

In sum, studies examining the first edition of the LASSI have yielded mixed results. Studies into the effects of teaching interventions on LASSI scores have yielded generally positive results. The studies in which students are taught specific skills for a whole semester (e.g., McKeachie et al., 1985; Nist et al., 1990) resulted in significant increases in LASSI scores, whereas studies with the low dosage or no interventions (e.g., Landrum & Chastain, 1998; Schumacker et al., 1995) did not result in systematic positive changes in LASSI scores. The Nist et al. study also suggested that LASSI scores might not be useful as predictors of academic achievement in at-risk college students who are most likely to need intervention in college. However, the study by Albaili (1997) did find that LASSI scores from a translated version differed between high and low achieving students and yielded a significant classification function, but this study did not report effect sizes for the differences, nor the percentages of students correctly classified.

Reliability of LASSI scores. Another set of studies on the first edition LASSI focused on the psychometric properties of the instrument's scores. Olejnik and Nist (1992) examined the internal consistency of LASSI scores in a sample of first-year college students. The reliability estimates that were similar to the estimates reported in LASSI manual (Weinstein, 1987), with a median alpha of .79. Deming et al. (1994) and Melancon (2002) also examined the reliability of LASSI scores and reported similar findings (see Table 1).

Structural validity of LASSI scores. Despite the fact that the LASSI is described as multidimensional and the individual subscale scores are interpreted for intervention, there are few studies examining the structural validity of the instrument's scores, and only one to date that has accomplished item-level scaling on the college-level instrument. Given the LASSI's 10 rationally-determined subscales, Melancon (2002) attempted to extract 10 factors using varimax rotation and principal components analysis, but the structure was not supported. Similarly, Olejnik and Nist (1992) examined the factor structure of the LASSI subscales, rather than the LASSI items. They hypothesized

Table 1

Internal Consistency Reliability Coefficient Alphas for LASSI Scores

LASSI Subscales	Weinstein (1987)	Olejnick & Nist (1992)	Deming et al. (1994)	Melancon (2002)	Current Study
Attitude	.72	.70	.72	.72	.77
Motivation	.81	.83	.76	.81	.83
Time Management	.86	.83	.79	.84	.86
Anxiety	.81	.82	.76	.84	.85
Concentration	.84	.85	.77	.86	.86
Information Processing	.83	.77	.78	.80	.82
Selecting Main Ideas	.74	.72	.68	.77	.77
Study Aids	.68	.65	.40	.67	.57
Self-Testing	.75	.80	.71	.74	.78
Test Strategies	.83	.79	.75	.80	.84

Note. LASSI = Learning and Study Strategies Inventory (Weinstein, Schulte, & Palmer, 1987).

that LASSI subscales might make up two factors representing *skill* and *will* as described by Paris, Lipson, and Wilcox (1983), or three factors described by Rohwer and Thomas (1987): selective allocation activities, processing activities, and cognitive monitoring activities. Olejnik and Nist extracted two- and three-factor solutions of subscale scores using principal axis extraction with promax rotation. Based on a cut-score of .40, the three-factor solution proved more viable.

Olejnik and Nist (1992) used confirmatory factor analysis in a second sample to test the two solutions, and found greater support for the three-factor solution. Factor 1 - Effort-Related Activities consisted of three subscales (Motivation, Time Management, & Concentration). Factor 2 - Goal Orientation was made up of Information Processing, Study Aids, and Self Testing, and Factor 3 – Cognitive Activities was comprised of Anxiety, Main Ideas, and Test Strategies. Olejnik and Nist (1992) concluded that the latent structure of LASSI subscales was similar to the theoretical constructs suggested by Rohwer and Thomas (1987).

Item factoring of the 76-item high school version of the LASSI (LASSI-HS; Weinstein & Palmer, 1990) was reported by Olivárez and Tallent-Runnels (1994). These researchers reported that "the individual items did not appear to cluster in their intended subscales" (Olivárez & Tallent-Runnels, p. 251) in either of the two samples used. Consequently, they examined the factor structure of the subscale scores, and they reported a three-factor solution similar to the one reported by Olejnik and Nist (1992). The same three-factor structure based on LASSI subscale scores was also reported in two Norwegian samples using a translated version of the LASSI (Olaussen & Bråten, 1998) and, more recently, in two American samples (Obiekwe, 2000).

The Present Study

The LASSI is a commercially available instrument that is used in colleges and universities to assess student difficulties in learning, to suggest areas for intervention, and to evaluate the effectiveness of interventions. Given the importance that is placed in LASSI scores (e.g., Robertson, 1994), establishing the reliability and validity of these scores continues to be important for educators and educational researchers. Five questions were posed in the current study: (a) What are the reliability estimates of LASSI scores based on published subscales, (b) What is the structural validity of LASSI item scores, (c) What is the structural validity of LASSI subscale scores, (d) How much variance in student grades do LASSI scores account for, and (e) Do LASSI scores of high and low achievers differ meaningfully?

Method

Participants

The participants consisted of 255 college students (25% male) who were taking an undergraduate educational psychology course at a major research university in the northeastern United States. They ranged in age from 17 to 47 ($M = 20$, $SD = 2.9$) and 94% were White. Students' semester standings ranged from first to 11th with a mean of 3.8 (SD = 1.5). The mean end of course score for students in the class was 84.3% ($SD = 11.7$).

Measure

Students completed the LASSI first edition (Weinstein et al., 1987), a 77-item, normative self-report questionnaire normed on 800 entering freshmen at a university in the southern US (Weinstein, 1987). Scales of Attitude, Motivation, Time Management, Anxiety, Concentration, Information Processing, Study Aids, Self-Testing, and Test Strategies scales are comprised of eight items each, whereas Selecting Main Ideas consists of five items. Internal

consistency estimates of scores range from .68 to .86, and test-retest reliability estimates range from .72 to .85 (Weinstein). Students rate themselves on each item using a 5-point Likert scale. After reversing the scoring on negatively worded items, raw scores are summed to compute the 10 subscale scores which are then plotted on a chart with percentile rankings to see how students compare to their peers. Higher scores indicate more behaviors and attitudes associated with academic success.

Procedure

Students in this study were volunteers who completed the questionnaire during one of their educational psychology class sessions during the semester. The students also completed a demographic form. Students' grades in the course were collected at the end of the semester and matched to the unique identifiers chosen by the students.

Results

Mean scores on the 77 individual items ranged from 2.5 to 4.8, and only one item had extreme skew and kurtosis scores. Scores on the 10 LASSI subscales were calculated by summing the items and dividing by the number of items to keep all subscales on the 1 – 5 scale. These scores ranged from 2.96 to 4.01. Intercorrelations among the 10 subscales were positive and ranged from .03 (between Study Aids and Anxiety) to .71 (between Test Strategies and Selecting Main Ideas; see Table 2). Three of the 10 subscales were correlated with final grades: Motivation (.37), Anxiety (.32), and Test Strategies (.39).

Reliability Estimates

Internal consistency estimates (Cronbach's alpha) were calculated for the 10 subscales and revealed a median coefficient of .83 (see Table 1 comparing these coefficients to those from prior studies). Reliability estimates for the subscales are generally in the .7 to .8 range across studies, with the exception of scores on Selecting Main Ideas, which consistently resulted in the lower internal consistency estimates.

Factor Analyses of LASSI Scores

As indicated previously, the literature contains only one factor analysis of LASSI item scores in college students from a Southern university (Melancon, 2002); however, no alternative factor solution was presented. As a result, exploratory factor analysis was applied in our study. Principal axis extraction was used and both orthogonal and oblique rotations were examined. The Kaiser (1974) measure of sampling adequacy for the correlation matrix was .88, and Bartlett's test of sphericity was significant, X^2 (2926) = 9735.21, $p < .001$, indicating that the matrix was factorable. The number of factors to extract was decided using multiple criteria (Carroll, 1985; Comrey, 1978, 1988; Thompson & Daniel, 1996). The eigenvalue greater than one rule suggested 19 factors and the scree test suggested 6 factors.

However, parallel analysis, using both the tables by Lautenschlager (1989) and computer software by Watkins (2000), suggested 4 factors. Communality estimates varied widely (i.e., .3 to .8), and the variable to factor ratio was approximately 20:3, suggesting that a sample size of 200 was

Table 2

Intercorrelations between LASSI Subscale Scores and Participants' Final Course Grade

	LASSI Subscale									
LASSI Subscale	Mot	Time	Anx	Con	Info	Main	Study	Self	Test	Grade
Att	.65	.46	.38	.64	.25	.43	.30	.33	.63	.19
Mot		.65	.25	.63	.34	.41	.40	.44	.56	.37
Time			.24	.66	.32	.36	.41	.54	.47	.27
Anx				.46	.11	.56	.03	.09	.65	.32
Con					.34	.58	.33	.44	.68	.25
Info						.34	.53	.65	.29	.07
Main							.21	.31	.71	.28
Study								.60	.18	.15
Self									.29	.08
Test										.39

Note. LASSI = Learning and Study Strategies Inventory. Att = Attitude, Mot = Motivation, Time = Time Management, Anx = Anxiety, Con = Concentration, Info = Information Processing, Main = Selecting Main Ideas, Study = Study Aids, Self = Self-Testing, Test = Test Strategies, Grade = participants' final grade in course. All correlations above .18 are significant at the .001 level.

sufficient for a convergent and admissible solution (MacCallum, Widaman, Zhang, & Hong, 1999). Structure coefficients of .35 (at least 12% of shared variance) were used for salience.

Multiple solutions were examined based on guidelines for factor extraction (Comrey, 1978; Floyd & Widaman, 1995). Using the scree test as the upper limit and parallel analysis, which is typically more accurate (Thompson & Daniel, 1996), as the lower limit, four-, five-, and six-factor solutions were examined. Two additional solutions were also examined: a 10-factor solution in keeping with the number of subscales scored, and a three-factor solution since factor analyses of both the LASSI and the LASSI-HS subscale scores resulted in similar three-factor structures.

Three-factor item solution. The three-factor solution based on item scores accounted for 34.2% of the variance and was interpretable. Given the large number of items on the LASSI (i.e., 77), the factor structure of the items is not included in this manuscript, but can be obtained from the authors. Table 3 contains a summary of the numbers of items that load on each of the three factors from both the orthogonal and the oblique rotations. In the orthogonal solution, four items cross-loaded and seven did not achieve a loading of at least .35 on any factor; in the oblique solution, the figures were zero and 10 respectively. Study Aids' items were the only ones where the majority did not load on a single factor—four of the eight Study Aids items had non-salient loadings. The three factors were identical to the ones labeled Effort-Related Activities, Goal Orientation, and Cognitive Activities in previous studies (Obiekwe, 2000; Olaussen & Bråten, 1998; Olejnik & Nist, 1992; Olivárez & Tallent-Runnels, 1994). However, we felt that neither these labels nor the Rohwer and Thomas (1987) constructs adequately described the factors, and opted to use labels that were more descriptive of factor content. Factor 1 consisted of Motivation, Time Management, Concentration, and Attitude items—4 of the 5 affective subscales—and was labeled Task Orientation. Factor 2, made up of Anxiety, Test Strategies, and Selecting Main Ideas, was labeled Test Knowledge to reflect the increased importance of these variables in assessment situations. Factor 3, which consisted of Information Processing, Self-Testing, and Study Aids, was labeled Comprehension, as each of these subscales is concerned with students' attempts to understand information. Subscale intercorrelations from the oblique rotation were also low: Factor 1-Factor 2 = .37, Factor 1-Factor 3 = .32, and Factor 2-Factor 3 = .11.

Three-factor subscale solution. The correlation matrix for the subscales was factorable: Kaiser's (1974) measure of sampling adequacy was .86, and Bartlett's test of sphericity was significant, $X^2(45) = 1304.15, p < .001$. The three-factor solution based on subscales accounted for 65% of the variance in subscale scores. Both the orthogonal and oblique rotation produced similar structures. Three factors were

Table 3

Number of Salient and Non-Salient Loadings for LASSI Three-Factor Item Solution[a]

LASSI Subscales	Factor 1 Task Orientation	Factor 2 Test Knowledge	Factor 3 Comprehension	Non Salient
Motivation	8(8)	0(0)	0(0)	0(0)
Time Management	8(8)	0(0)	1(0)	0(0)
Concentration	7(8)	1(0)	0(0)	0(0)
Attitude	5(5)	1(0)	0(1)	2(2)
Anxiety	0(0)	8(8)	0(0)	0(0)
Test Strategies	3(0)	8(7)	0(0)	0(1)
Selecting Main Ideas	0(0)	4(4)	1(1)	0(0)
Information Processing	0(0)	0(0)	8(8)	0(0)
Self-Testing	1(0)	0(0)	7(5)	1(3)
Study Aids	0(0)	0(0)	4(4)	4(4)

Note. LASSI = Learning and Study Strategies Inventory. All subscales have 8 items, except Selecting Main Ideas, which has 5 items.

[a]The first set of numbers in the columns are from the varimax rotation, and the second set of numbers are from the pattern matrix of the oblimin rotation.

Table 4

Pattern Coefficients for LASSI Subscales from Three-Factor Subscale Solution

LASSI Subscales	Factor 1 Task Orientation	Factor 2 Comprehension	Factor 3 Test Knowledge
Motivation	.88	.02	-.06
Time Management	.66	.19	-.02
Attitude	.65	-.06	.22
Concentration	.59	.08	.32
Information Processing	-.12	.80	.13
Self-Testing	.13	.80	-.03
Study Aids	.18	.62	-.12
Anxiety	-.01	-.07	.78
Test Strategies	.31	-.01	.72
Selecting Main Ideas	.04	.19	.70
Eigenvalues	4.91	1.75	.89
Percent Variance	45.80	13.84	5.24
Construct Reliability	.86	.79	.78

Note. LASSI = Learning and Study Strategies Inventory.

Table 5

Variance in Final Grade in Course Accounted for by LASSI Scores

Predictor Variables	*B*	β	sr^2	Adj. R^2
Attitude	-5.25	-.27*	-.20	
Motivation	6.73	.38**	.27	
Time Management	1.51	.10	.08	
Anxiety	2.59	.17	.14	
Concentration	-2.05	-.12	-.08	
Information Processing	-1.06	-.06	-.05	
Selecting Main Ideas	-.20	-.01	-.01	
Study Aids	2.87	.13	.12	
Self-Testing	-3.04	-.16	-.12	
Test Strategies	6.05	.33*	.20	.25**

Note. LASSI = Learning and Study Strategies Inventory. sr^2 = squared semipartial correlation.
*$p < .01$. **$p < .001$.

assigned the same names as those in the item-based solution (see Table 4): Factor 1 was labeled Task Orientation, Factor 2 was labeled Comprehension, and Factor 3 was labeled Test Knowledge.

Regression Analysis

A regression equation was calculated using final grade in the course as the dependent variable and the 10 LASSI subscale scores as predictors. The equation was significant with the LASSI subscales accounting for 25% of the variance in final scores (see Table 5). However, only three of the subscales contributed significantly to the equation: Attitude, Motivation, and Test Strategies.

Group Differences

Students were divided into three roughly equal groups based on final grade in class and the differences between the top group and the bottom group on the 10 LASSI subscales were examined using independent *t*-tests. The critical alpha was set at .005 to control for error rate. Seven of the 10 differences were significant (see Table 6), and effect sizes (Cohen's *d*) ranged from .14 to 1.02, with eight falling in medium to large range (Newton & Rudestam, 1999).

Discussion

In this study, we examined the internal consistency and structural validity of scores on the LASSI first edition, as well as differences between students who obtained the lowest and highest grades in a university course. Almost all (9 of the 10) of the LASSI subscale scores resulted in internal consistency estimates in moderate to high range. Moreover, this finding is consistent with those in the normative sample (Weinstein, 1987) and previous studies (Albaili, 1997; Deming et al., 1994; Melancon, 2002; Olaussen & Bråten, 1998; Olejnik & Nist, 1992). While it is important that scores be internally consistent, it is important to remember that reliable scores are not necessarily evidence of validity (Goodwin & Goodwin, 1999). The high internal consistency estimates tell us that individuals' responses on the items that are assigned to subscales tend to be similar.

One criterion for robustness of research findings is replicability, and the results of this study contribute to that criterion with regard to the structural validity of LASSI scores. The three factor structure found here replicates those previously found with college students at universities in the US (Olejnik & Nist, 1992; this study) and in Norway (Olaussen & Bråten, 1998) and with high school students in the US (Olivárez & Tallent-Runnels, 1994). This study is the first to report an interpretable factor structure for LASSI items; however, the principle of parsimony, as well as the increased error inherent in factoring items (Bernstein & Teng, 1989), suggests that the LASSI structure should be determined at the subscale rather than the item level. The studies to date lend credence to that contention.

LASSI subscale scores accounted for 25% of the variance in final course grade in this study, with three

Table 6

Mean Scores on LASSI Subscales by Final Grade in Course

LASSI Subscale	Total Sample		Bottom Third on Final Grade		Upper Third on Final Grade		Cohen's *d*
	M	*SD*	*M*	*SD*	*M*	*SD*	
Attitude*	4.01	.59	3.83	.59	4.18	.59	.59
Motivation*	3.99	.66	3.66	.59	4.33	.57	1.02
Time Management*	2.96	.79	2.72	.71	3.29	.85	.72
Anxiety*	3.18	.77	2.88	.75	3.52	.72	.83
Concentration*	3.27	.70	3.06	.61	3.51	.70	.64
Information Processing	3.40	.63	3.39	.62	3.48	.67	.14
Selecting Main Ideas*	3.70	.65	3.48	.62	3.91	.62	.66
Study Aids	3.13	.54	3.03	.55	3.25	.50	.41
Self-Testing	2.96	.62	2.91	.59	3.07	.65	.26
Test Strategies*	3.67	.63	3.37	.54	3.97	.57	.95

Note. LASSI = Learning and Study Strategies Inventory.
*$p < .001$.

subscales contributing significantly. This finding is comparable to the 20% reported by Nist et al. (1990) in another group of college students, and the canonical correlation of .55 (30% variance) reported by Albaili (1997) in college students in the United Arab Emirates. However, several researchers have reported that LASSI scores do not predict additional variance in academic achievement of college freshmen beyond the predictions of background variables and variables used for admission (e.g., SATs, high school GPA; Rugsaken, Robertson, & Jones, 1998) or those found to be predictive of college grades (e.g., SATs, undergraduate class rank, undergraduate GPA, and demographics; Young, 1994). As these variables were not included in the current investigation, their inclusion in subsequent investigations would be beneficial.

Findings on the meaningfulness of group differences are also not fully resolved by this study. Students who did least well in the course obtained significantly lower scores on seven subscales than students who performed in the top third of the class, and the differences were substantial. These findings are comparable to those of Albaili (1997) and Nist et al. (1990). However, the students in both this study and Albaili's were regular education students. Some studies of previously identified at-risk students have yielded equivocal results (Deming et al. 1994). In fact, in a very recent study, Hewlett, Boonstra, Bell, and Zumbo (2000) found that, as in previous studies, students who had substantial deficits in reading comprehension and vocabulary and were in need of remediation were misclassified as average or above-average readers by the LASSI.

This study had several limitations. First, the students were all volunteers. Second, it is not clear if any of them were identified with learning difficulties, thus limiting the generalizability of the findings. Third, the only external criterion used was final grade in the course that participants were taking. This single-item variable may not represent an adequate range of student performance, and may have resulted in correlations that were attenuated. Finally, although study skills were covered as a topic in the class, there was no attempt to intervene specifically with students, nor did we know if some students in the class were involved in some specific intervention elsewhere on campus.

Conclusion

The results of the current study indicate that LASSI first edition subscale scores result in moderate to high reliability estimates that are best represented by three factors, rather than the 10 subscales on which the instrument is scored. Although LASSI scores do predict variance in college grades, other studies indicate that the prediction may disappear when other predictors are included, and this is an area where further study is warranted. Consequently, LASSI scores may not be valid for the students who are most likely to be administered the LASSI, that is, students who are in serious need of remediation upon entry to college. The newly revised LASSI second edition (Weinstein et al., 2002) retains and similarly embraces the interpretation of the 10 rationally derived scales of the first edition. As this is a revised instrument, further investigation is merited; however, until the 10 scales are supported via factor analytic investigation, we must agree with Melancon (2002), who suggested that "an alternative scoring strategy may be required for LASSI data" (p. 1026). The three-factor structure which has been supported in factoring both items and subscales for the LASSI first edition can provide the foundation for revisions to the second edition, and the three factors should also be used as the starting point for developing effective remediation approaches to address identified study strategy deficits.

Finally, the current evidence of reliability and structural validity of LASSI scores are only a part of the construct validity evidence needed for a measure (Benson, 1998). Other evidence must be gathered using a variety of tools, including multitrait-multimethod analyses, experimental manipulation, and structural equation modeling (Benson). The extant literature suggests that more studies in the area of group differentiation and predictive validity are also warranted, using the three-factors rather than the 10 subscales. In closing, we echo Mealey's (1988) caution, and note that both the first and second editions of the LASSI should be considered experimental measures and not used for high-stakes decision making.

References

Albaili, M. A. (1997). Differences among low-, average-, and high-achieving college students on learning and study strategies. *Educational Psychology, 17,* 171-177.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-22.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105,* 467-477.

Carroll, J. B. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence: Volume 1, research methodology.* Norwood, NJ: Ablex Publishing Corporation.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46,* 648-659.

Comrey, A. L. (1988). Factor analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56,* 754-761.

Deming, M. P., Valeri-Gold, M., & Idleman, L. S. (1994). The reliability and validity of the Learning and Study Strategies Inventory (LASSI) with college developmental students. *Research and Instruction, 33,* 309-318.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286-299.

Goodwin, L. D., & Goodwin, W. L. (1999). Measurement myths

and misconceptions. *School Psychology Quarterly, 14,* 408-427.

Hewlett, M. G., Boonstra, J., Bell. J. H., & Zumbo, B. D. (2000). Can LASSI score profiles help identify postsecondary students with underlying reading problems? *Journal of College Reading and Learning, 30,* 135-143.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31-36.

Landrum, R. E., & Chastain, G. (1998). Demonstrating tutoring effectiveness within a one-semester course. *Journal of College Student Development, 39,* 502-506.

Lautenschlager, G. J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research, 24,* 365-395.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84-99.

McDermott, P. A. (1999). National scales of differential learning behaviors among American children and adolescents. *School Psychology Review, 28,* 280-291.

McKeachie, W. J., Pintrich, P. R., & Lin, Y. (1985). Teaching learning skills. *Educational Psychologist, 20,* 153-160.

McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology, 67,* 198-203.

Mealey, D. L. (1988). Learning and Study Strategies Inventory (LASSI) test review. *Journal of Reading, 31,* 382-385.

Melancon, J. G. (2002). Reliability, structure, and correlates of Learning and Study Strategies Inventory scores. *Educational and Psychological Measurement, 62,* 1020-1027.

Newton, R.R., & Rudestam, K.E. (1999). *Your statistical consultant.* Thousand Oak, CA: Sage Publications.

Nist, S. L., Mealey, D. L, Simpson, M. L., & Kroc, R. (1990). Measuring the affective and cognitive growth of regularly admitted and developmental studies students using the Learning and Study Strategies Inventory (LASSI). *Reading Research and Instruction, 30,* 44-49.

Obiekwe, J. C. (2000, April). *The latent structure of the Learning and Study Strategies Inventory (LASSI). A comparative analysis.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED442873)

Olaussen, B. S., & Bråten, I. (1998). Identifying latent variables measured by the Learning and Study Strategies Inventory (LASSI) in Norwegian college students. *Journal of Experimental Education, 67,* 82-96.

Olejnik, S., & Nist, S. L. (1992). Identifying latent variables measured by the Learning and Study Strategies Inventory (LASSI). *Journal of Experimental Education, 60,* 151-159.

Olivárez, A., Jr., & Tallent-Runnels, M. K. (1994). Psychometric properties of the Learning and Study Strategies Inventory-High School version. *Journal of Experimental Education, 62,* 243-257.

1 in 11 Freshmen Reports Having a Disability, Survey Finds (2000, February 4). [Electronic version]. *The Chronicle of Higher Education,* p. A53.

Paris, S. G., Lipson, M. Y., & Wilcox, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8,* 293-316.

Pintrich, P. R., McKeachie, W. J., & Lin, Y. (1987). Teaching a course in learning to learn. *Teaching of Psychology, 14,* 153-160.

Reynolds, W. M. (1979). Development and validation of a scale to measure learning-related classroom behaviors. *Educational and Psychological Measurement, 39,* 1011-1018.

Robertson, J. A. (1994). The Learning and Study Strategies Inventory: A versatile assessment enhances student success. *Research and Teaching in Developmental Education, 11,* 15-20.

Rohwer, W. D., Jr., & Thomas, J. W. (1987). The role of mnemonic strategies in study effectiveness. In M. A. McDaniel & M. Pressley (Eds.), *Imaginal and mnemonic processes* (pp.428-450). New York: Springer-Verlag.

Rugsaken, K. T., Robertson, J. A., & Jones, J. A. (1998). Using the Learning and Study Strategies Inventory scores as additional predictors of student academic performance. *NACADA Journal, 18,* 20-26.

Schumacker, R. E., Sayler, M., & Bembry, K. L. (1995). Identifying at-risk gifted students in an early college entrance program. *Roeper Review, 18,* 126-129.

Stott, D. H., Green, L. F., & Francis, J. M. (1983). Learning style and school attainment. *Human Learning, 2,* 61-75.

Stott, D. H., McDermott, P. A., Green, L. F., & Francis, J. M. (1988). *Learning Behaviors Scale and Study of Children's Learning Behaviors research edition manual.* San Antonio, TX: The Psychological Corporation.

Thompson, B., & Daniel, L G., (1996)). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56,* 197-208.

Watkins, M. W. (2000). Mac Parallel Analysis [Computer software]. State College, PA: Author.

Weinstein, C. E. (1982). *Learning and Study Strategies Inventory.* Unpublished manuscript. University of Texas at Austin.

Weinstein, C. E. (1987). *Learning and Study Strategies Inventory (LASSI): User's manual.* Clearwater, FL: H & H Publishing.

Weinstein, C. E., & Palmer, D. R. (1990). *Learning and Study Strategies Inventory (LASSI)-High School version: User's manual.* Clearwater, FL: H & H Publishing.

Weinstein, C. E., & Palmer, D. R. (2002). *Learning and Study Strategies Inventory – Second Edition User's Manual.* Clearwater, FL: H & H Publishing.

Weinstein, C. E., Palmer, D. R., & Shulte, A. C. (2002). *Learning and Study Strategies Inventory – Second Edition.* Clearwater, FL: H & H Publishing.

Weinstein, C. E., Schulte, A. C., & Palmer, D. P. (1987). *Learning and Study Strategies Inventory (LASSI).* Clearwater, FL: H & H Publishing.

Young, J. W. (1994). Differential prediction of college grades by gender and by ethnicity: A replication study. *Educational and Psychological Measurement, 54,* 1022-1029.

Barbara A. Schaefer received her Ph.D. in School, Community and Clinical Child Psychology from the University of Pennsylvania and she is currently an Assistant Professor in the Educational and School Psychology and Special Education Department at Pennsylvania State University. Her research interests are in the areas of learning behaviors, psychometrics, and assessment and intervention to assist in improving academic outcomes.

Frank C. Worrell, Ph. D. is an Associate Professor and the Director of UC Berkeley's APA-accredited doctoral school psychology program, and faculty consultant to the Academic Talent Development Program. Dr. Worrell teaches courses in consultation, dropout prevention, and gifted and talented education. His research interests include education in English-speaking Caribbean, teacher effectiveness, scale development, and psychosocial factors in the education of African American, at-risk, and gifted and talented youth.

Author Note

The authors appreciate the assistance of Dr. Rayne Sperling with data collection.

Correspondence concerning this article should be addressed to Barbara A. Schaefer, ESPSE: School Psychology Program, 227 CEDAR Building, University Park, PA 16802-3108. E-mail: bas19@psu.edu.