The Journal of Research in Education

Fall 2013 Volume 2

TABLE OF CONTENTS

Against Conventional Wisdom: Factors Influencing Hispanic Students' Reading Achievement

Author Info:

Mr. Percell is an Instructional Assistant Professor at Illinois State University. Ms. Kristina Kaufman is a doctoral student in Education at Illinois State University.

Correspondence regarding this manuscript should be addressed to Jay C. Percell, College of Education, Campus Box 5300, Normal IL 61790. Ph: (309) 438-3836. Email: jpercel@ilstu.edu

Abstract

The researchers performed a variable analysis of the 2002 Educational Longitudinal Study data investigating factors that influence students' reading scores on standardized tests. Hispanic and non-Hispanic Scores were analyzed and controlling variables were compared to determine the effect of each on both populations. Certain variables commonly thought to positively influence students' reading scores, such as family background, proved less statistically significant among the Hispanic population. Additionally, other variables usually associated with lower reading scores, such as urbanicity, were not. Implications of these findings are discussed and educators are encouraged to rethink variables that impact reading achievement among Hispanic students.

Keywords: Standardized Reading Test Scores, Hispanic Students

Introduction

According to Greenleaf et al. (2011) "our democracy and future economic well-being depend on a literate populace, capable of fully participating in the demands of the 21st century" (p. 648). Those demands have been well documented (Carnegie Council on Advancing Adolescent Literacy, 2010; Elrod, 2010; Futrell, 2010; Partnership for 21st Century Skills, 2011; Silva, 2008) and are reoccurring themes among education discourse which include the acquisition of a variety of "literacies" including information, media and ICT literacy as well as communication skills (Fleischman, Hopstock, Pelczar & Shelley, 2011; Partnership for 21st Century Skills, 2011). Personal and professional success begins with education and language development to create and communicate concepts and ideas. Thus, reading achievement and literacy practices in varied contexts are important foci for educators.

This study describes the analysis of data of over 15,000 reading assessment scores of 10th grade students in the United States. Data taken from the 2002 Educational Longitudinal Study (ELS) was analyzed to investigate and understand the variables that may influence reaching achievement among the Hispanic student population in this group as compared with their non-Hispanic peers.

Review of Literature

The ability to read and write and communicate with others is foundational to fulfilling academic, personal and professional goals. A reoccurring theme of student success is in the strength of their reading skills (Cooper, Kiger, Robinson & Slansky, 2012). Caspe (2009) cites literacy as a "critical developmental accomplishment" for children (p. 306). There is even more importance placed on the attainment of solid literacy skills with the influx of technology (Joseph & Schisler, 2009). Schools across the nation are faced with preparing students for successful performance on standardized tests that assess students' reading skills.

Sound literacy skills are crucial for not only the assessment laden school environment but in other aspects of a students' education and in their eventual career path (Lee, Olszewski-Kubilius, & Peternel, 2010). Being able to communicate effectively and clearly with others is vital in today's world and in one's ability to also receive and interpret communications as well. One research supported way in which students can improve reading skills is through – reading. "The amount of reading students engage in has been shown to be a strong predictor of academic achievement" (Mucherah & Yoder, 2008, p. 214). Just how teachers position and assign reading can play a role in students' reading, but also external factors such as their exposure to print-rich environments, support and communication with and from family, family income to provide resources, and interest level. But the exposure to literacy practices and reading both in and out of school is vital to student success in school and in the 21st century information society (Aydin, Erdagf, & Tas, 2011; Wamba, 2011; Greenleaf et al., 2011, Snipes & Horwitz, 2008).

Wamba (2011) ascertains the concern over children and reading in the following: **"**Reading and writing are passports to achievement in many other curricular areas, and literacy education plays an important role in moving people out of poverty toward greater self-sufficiency post-graduation. Schools and home environments share responsibility for literacy skill development" (p. 8).

A student does not come to the classroom a blank slate, however, but is riddled with experiences that shape the students' approach to school, ability to perform, and comprehension. Research indicates that parental involvement in a students' schooling can greatly impact how the child succeeds (Auerbach, 1997; Gaitan, 2012; Ortiz, 2004).

Because reading and literacy are important to understandings in the field of education, this study will focus on the analysis of data relating to reading achievement scores of Hispanic students and aspects of personal background that may affect a students' score. It is important to consider the students that make up the population in our nation's classrooms. The U.S. is on a trajectory to continue to become increasingly diverse (Ball & Tyson, 2011; Boske & Benavente-McEnery, 2010). Hispanics, especially, are the group of individuals that are the fastest growing subset of the U.S. population (Hemphill & Vanneman, 2011; Humes, Jones, & Ramirez 2011; Kober, 2010). Moreover, approximately 12% of people age five and over in the U.S. are Hispanic (Robinson, 2008). Some research indicates that there is an achievement gap between White students and Hispanic students (Robinson, 2008; Rojas-LeBouef & Slate, 2012; Lopez et. al, 2007). Kober (2010) cites that by eighth grade, Hispanic students are only 58% proficient in reading by 8th grade and only 56% of Hispanic high school students are proficient in reading compared to 81% and 78% respectively for Whites.

Achievement gaps can be attributed to a variety of factors. Certainly, the dominance of monolinguistic, white female teachers in education may have play a role in the ability to connect with diverse groups of students (Ball & Tyson, 2011; Darling-Hammond & Bransford, 2005). Some studies have suggested that variables of income, parental education and occupation, and immigration status may play a role in Hispanic students' reading development (Grouws, 1992; Pond, 1999). Understanding the factors that hinder or support literacy development is important as research has indicated a link between literacy development and achievement later on in life (Billings, 2009; Dickinson & Tabors, 2002; Herbers et al., 2012).

Family involvement is one aspect that this study aims to look at more closely. Family may play an instrumental role in literacy development among children (Billings, 2009; Ortiz, 2004; Whitehurst et al., 1988). Moreover, a family's income can afford for additional opportunities for learning or be restricted by financial implications. Poverty is a great issue facing many students today. Berliner (2006) points out that poverty is the issue that is most plaguing student achievement and that students of urban minority and poor students are below that of their middle-class white peers. "A majority of school-age Latino children are economically disadvantaged. More than one-fourth (27%) come from families with incomes below the poverty level, and another 33% are near poor" (Kober, 2010, p. 3).

We examined the data in this study's sample of students to identify with or challenge this notion by analyzing Hispanic students' family structure, access to a computer in the home, urbanicity, and whether students think reading is fun or not as compared to all other non-Hispanic students in this 15,362 student dataset.

Methodology

A descriptive analysis approach was first applied to the 2002 ELS dataset regarding students' personal backgrounds and their standardized reading scores (see Table 1). This information prompted us to investigate specific variables and their individual significance upon students' reading scores. Furthermore, we were particularly interested in how Hispanic students' reading scores were affected by their personal backgrounds, especially given Hemphill and Vanneman's (2011) indication that Hispanics are the fastest growing population in the US.

Therefore, we recoded the race variable of the ELS dataset to distinguish between Hispanic and non-Hispanic students. Students identifying themselves as either "Hispanic, no race specified", or "Hispanic, race specified" were coded as "1". Students identifying as White, non-Hispanic, Black or African American, Asian, American Indian or multiracial, were coded as "0". All other possible entries or omissions in the category of race were coded as "missing data." After the recoding, we ran a frequency distribution to identify the number of students identifying to some degree as Hispanic (Yes = 1) was N=2,440, and the number of students identifying as non-Hispanic (No = 0) was N=12,922.

Next, we generated split form data in order to examine the sample populations side by side: Hispanics and all non-Hispanics. Once data was spilt, we then isolated variables and ran linear regressions to determine the significance of each variable. The first variable examined was whether or not students thought reading was fun. In order to examine how students' enjoyment of reading affected their standardized reading test scores, we ran a linear regression and displayed the results as a split form to analyze the difference of this variable between Hispanic students and non-Hispanic students. The second variable analyzed was whether or not students' families had a computer at home, and how analyzed how that variable affected reading scores. The third variable analyzed was students' family structure, and whether or not being raised in a "traditional" family (students living with both mother and father at home) had an impact on reading scores. The final variable examined was urbanicity, and to what degree living in an urban setting affected students' reading scores.

Finally, we tabulated all models and variables into a split form, multi-category regression analysis (see Table 6). This allowed us to examine the specific effect of each independent variable upon the dependant variable of students' standardized test scores in reading, while simultaneously controlling for all others. Because the data was split, it was easier to make a visual comparison between the two populations: Hispanic and non-Hispanic. Our hypotheses are as follows:

H_0 = Personal background has no impact on Hispanic students' standardized reading scores.

H_1 = Personal background does have an impact on Hispanic students' standardized reading scores.

When we indicate "personal background" in our hypotheses, we are acknowledging that there are several variables that may influence a student's reading score. The

variables of family construct, urbanicity, whether or not a student has a computer in their home and their preference for reading are variables of consideration in this study.

Findings

After recoding the race variable and generating split form data, the first variable that we examined in relation to reading standardized test scores was students' interest in reading. The p-values for the variable of whether students "Thinks reading is fun" are less than .05 for both Hispanics and non-Hispanics. This variable showed a strong statistical significance for non-Hispanic students based on the linear coefficient, with a Beta score of B=4.272 and a Sig. = .000 (see Table 2). However, this was not the case among the Hispanic population. The variance of Hispanic scores was considerably low, at $R^2 = .076$, indicating that reading enjoyment could only account for less than 1% increase in test scores. Furthermore, the Beta score was also low, at B= 1.448, verifying that this variable was not very statistically significant among the Hispanic population. It had been our assumption that an enjoyment of reading would likely lead to improved reading scores, as the non-Hispanic scores demonstrated. It was striking to note that this was indeed not the case among the Hispanic population. Hispanic students' reading scores were largely unaffected, regardless of whether the students enjoyed reading or not.

Beyond a pure enjoyment of reading, we were interested to see the how the prevalence of access to technology affected students' reading scores. We examined the significance of whether or not the family owned a computer, as it related to their reading scores. As shown in Table 3, the variable of reading scores had a strong statistical significance among both populations. Examining Hispanic students, this variable produced a Beta score of B= 4.839 and a Sig. = .000. Additionally, among the non-Hispanic groups, it was likewise significant, with a Sig. = .000 and a B= 7.264. Furthermore, the Confidence Interval at 95% was much higher for the variable of "Computer" ownership than it was for "Thinks reading is fun" at 5.943 in the upper bound as opposed to 2.428 for Hispanics, and 7.906 in the upper bound compared with 4.700 for non-Hispanics. Still, despite this variable's significance among Hispanics, this data implies that computer ownership still accounts for a greater increase in reading scores among the non-Hispanic population.

Our third control variable to analyze was "Family Structure" (see Table 4). This variable had been recoded to indicate students who lived with both father and mother (Yes = 1), as opposed to any other family living situation (No = 0). The regression analysis indicated this variable as having a statistical significance upon the reading scores of the non-Hispanic sample. Their Beta scores were high, with a B= 3.215 and a Sig = .000, although notably lower than the "Computer" variable. However, much to the surprise of the researchers, among the Hispanic group, the regression showed this variable as having no real statistical significance. Much like the "Thinks reading is fun" variable, living at home with both father and mother produced a low Beta score among Hispanics, B = .833 and a Sig = .074. Moreover, the variance was very low, with $R^2 = .043$, indicating that this variable can only account for 4.3% improvement in reading scores among this population.

Finally, the fourth variable examined related to students' standardized reading scores and

was the variable "Urban" (see Table 5). Once again, we recoded this variable as students living in an Urban setting (Yes = 1), versus those who did not (No = 0), and this variable produced the most diverse figures of any variable tested, as it was the only one to produce a negative association. For both the Hispanic and non-Hispanic groups, the Beta scores were negative, with B = -.131 and B= -.358, respectively. This likewise produced a negative variance, with an R^2 = -.007 among Hispanics and R^2 = -.016. This means, that for every unit increase in the "Urban" variable, the standard deviation of reading scores will decrease - a 35.8% decrease for non-Hispanics and 13.1% decrease for Hispanics.

Although the significance scores for both groups were higher, Sig = .083 for non-Hispanics and Sig = .776 for Hispanics, those scores are not true indicators of this variable's statistical significance, due to the fact that the association is negative. What was telling was that the t^2 score for the non-Hispanic group at -1.735 was closer to 2, a general rule of thumb for indications of significance (Vogt, 2007). However, the t^2 score for the Hispanic sample was -.284, nowhere near the generally accepted level of significance. Therefore, urbanicity, despite having a negative correlation with students' reading scores among the entire population, had no real statistical significance upon the reading scores of the Hispanic students which is telling. While the data has shown the average reading standardized test score of Hispanics to be less than non-Hispanics, something which Berliner (2006) highlights, urbinicity does not fully explain the lower performance of Hispanics as compared to non-Hispanics. In fact, in regard to standardized reading test scores, urbanicity affects Hispanic students less that it does non-Hispanic students.

Finally, all variable in this model were compiled in a multi-categorical regression analysis so as to display the adjusted R^2 value when controlling for all variables (see Table 6). It should once again be noted that do to the rather large sample size, even small variations in percentages demonstrate a strong significance, for instance, that 12% of non-Hispanic students reading scores can be directly attributed to these four variables in analysis is significant, especially considering that n = 12,922. Still, the Adjusted R^2 value for the Hispanic population, despite its size (n = 2,440), increased very little. Indeed, when controlling for these four variables, it seemed to flatten. These four variables, some of which were thought might have a significant impact on reading scores, ultimately were of no real significance to the Hispanic students. Since the Adjusted R^2 = 0.048 when controlling for these variables at p = 0.05, we must fail to reject the null. We accept that personal background, insofar as it is defined by these four variables, has no impact on Hispanic students' standardized reading scores in contrast to all other students.

Implications

Based on the data we analyzed, there are interesting implications that surface when examining multiple variables' effects on students reading scores. It is interesting to note that of all variables tested, whether or not the family owned a computer had the greatest significance (B= 7.263; B = 4.839). This would seem to suggest that the most important factor to increase a student's literacy development as related to their reading scores would be access to technology, particularly computers. There could be several reasons for this. Given the amount of educational software programs available through interactive devices like

computers or tablets, students who have access to them would likely have a decided advantage over students who did not (Norris, 2003). Additionally, when thinking of technology itself as a literacy, especially in light of 21st century skill development, an increased awareness in a technological literacy could have direct benefit to reading literacy. The significant impact of computers upon students reading skills, specifically among Hispanics, is consistent with Vassiliou's (2011) research with computer assisted software technology and ESL students' literacy rates in Miami, finding that students' scores increased considerably when using the technology. Furthermore, Taningco and Pachon (2008) also found that computer use both at home and in the classroom has a positive correlation on Latino students' scores in mathematics. Certainly, there is an argument to be made that students with access to technology and educationally rich software can improve their academic success.

Additionally, it was interesting to note that while the "Family Structure" variable and the "Urban" variable were very significant among the non-Hispanic population, both variables showed no real statistical significance among Hispanic students. The reasons behind this are baffling. Perhaps, because of the proliferation of an expanded nuclear family unit among many Hispanic families, with greater reliance on grandparent, and aunts and uncles than other American cultural groups (Hsueh-Fen, Lynn & Kyungeh, 2012) it could be that Hispanic students who do not live with both their mother and father do not suffer academically as a result, unlike many non-Hispanic students. Perhaps the strong family network that many Hispanic communities employ is able to overcome the absence of both a mother and father living at home together.

Additionally, the implications of urbanicity having no statistical significance upon the reading scores of Hispanic students are also interesting to note. This result might imply that a large percentage of Hispanics come from urban regions, and that this variable has very little impact upon students scores therefore their scores will be unaffected. However, another interpretation could be that overall Hispanic reading scores are lower than non-Hispanics to begin with, that regardless of whether students are urbanized or not makes little difference. Either way, this conclusion is troubling, especially given the negative association of the Urban variable, and the fact that it still had little impact on the Hispanic students' scores.

Regardless, when weighed together, especially in light of the researchers own expectations and hypothesis in conducting this study, there is a serious implication for the need to critically examine our own biases and assumptions, both as researchers and educators. To assume that one set of variables will have the same impact on different subgroups of students, especially students hailing from different cultural backgrounds, is simply erroneous, as demonstrated by the relatively static Adjusted R^2 value among Hispanic students in our multi-categorical regression. One interpretation of these findings, as the present achievement gap would suggest, is that the American education has failed to completely acculturate Hispanic students to measurable levels of success and greater attention is needed to support these students in the coming years given population trends.

Conclusions

United Nations Secretary General Khafi Annan said that literacy is a bridge from misery to hope (Annan, 1997). In light of that truth, we have analyzed several

independent variables relating to students' standardized test scores in reading. While literacy and reading are crucial skills, as they are foundational for all learning among all students, we specifically focused on the variables affecting the reading scores from the Hispanic population in the 2002 ELS dataset. Not only are Hispanics the fastest growing ethnic group in the United States (Hemphill and Vanneman, 2011), but there also exists a stark achievement gap in reading and literacy scores between Hispanic and non-Hispanic students in this country (Poulsen, Hastings, and Allbritton, 2007).

Our research has revealed that some factors that impact students' reading scores such as traditional family structures and urbanicity, do not have the same statistical significance with Hispanic students. While it has been evidenced that access to computers and technology has a significant impact on all students' reading skills, further research is needed in order to find additional variables which impact reading scores specifically among Hispanic students. Kober (2010) noted that it is critical that Hispanic students are prepared for college, careers and civic participation since they are the population who will "shape the nation" (p.1). Therefore, as educators, we must do all we can to equip this essential ethnic population with essential skills of literacy.

References

Annan, K. (1997) *Secretary-general stresses need for political will and resources to meet challenge of fight against illiteracy*. Press Release SG/SM/6316 OBV/9. Retrieved from http://www.un.org/News/Press/docs/1997/19970904.SGSM6316.html

Auerbach, E. (1997). Reading between the lines. In D. Taylor (Ed.), *Many families, many literacies* (pp.71-81). Portsmouth, NH: Heinemann.

Aydin, A. Erdagf, C. & Tas, N. (2011). A comparative evaluation of pisa 2003-2006 results in reading literacy skills: An example of top-five OECD countries and Turkey. *Educational Sciences: Theory and Practice, 11*(2) 665-673.

Ball, A. & Tyson, C. A. (Eds.). (2011). *Studying diversity in teacher education*. Lanham, MD: Rowan & Littlefield.

Billings, E. (2009). Prescriptions to read: Early literacy promotion outside the classroom. *Literacy Teaching and Learning, 13*(1/2), 81-101.

Berliner, D. (2006). Our impoverished view of educational research. *Teachers College Record, 108*(6), 949-995.

Boske, C., & Benavente-McEnery, L. (2010). Taking it to the streets: A new line of inquiry for school communities. *Journal of School Leadership, 20*(3), 369-398.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York, NY: Carnegie Corporation of New York.

Caspe, M. (2009). Low-income Latino mothers' booksharing styles and children's emergent literacy development. *Early Childhood Research Quarterly*, *24*(3), 306-324.

Cooper, J., Keiger, N., Robinson, M., & SLansky, J. (2012). *Literacy: Helping students construct meaning.* Belmont, CA: Wadsworth.

Darling-Hammond, L., & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.

Dickinson, D. K., & Tabors, P. O. (2002). Fostering language and literacy in classrooms and homes. *Young Children*, *57*(2), 10–18.

Elrod, S. (2010). Project kaleidoscope 2.0: Leadership for twenty-first century STEM education. *Liberal Education*, *96*(4), 24-33.

Fleischman, H., Hopstock, P., Pelczar, M., & Shelley, B. (2011). Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context. NCES 2011-004. *National Center for Education Statistics*, 72pp.

Futrell, M. (2010). Transforming teacher education to reform America's p-20 education system. *Journal of Teacher Education*, *61*(5), 432-440.

Gaitan, C. (2012). Culture, literacy, and power in family-community-school-relationships. *Theory Into Practice, 54*(4), 305-311.

Greenleaf, C., Litman, C., Hanson, T., Boscardin, C., Herman, J., Schneider, S., Madden, S., & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, *48*(3), 647-717.

Grouws, D. (1992). *Handbook for research on mathematics teaching and learning*. New York: MacMillan.

Hemphill, F., & Vanneman, A. (2011*). Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress*. Washington, D.C.: National Center for Education Statistics.

Herbers, J., Cutuli, J., Supkoff, L., Heistad, D., Chan, C., Hinz, E., & Masten, A. (2012). Early reading skills and academic achievement trajectories of students facing poverty, homelessness, and high residential mobility. *Educational Researcher, 41*(9), 366-374.

Hsueh-Fen, S., Lynn, M., & Kyungeh, A. (2012). Development of the Mexican American family loyalty toward elderly relatives scale. *Journal of Theory Construction & Testing, 16*(2), 38-44.

Humes, K., Jones, N., & Ramirez, R. (2011). Overview of race and Hispanic origin: 2010. *U.S. Census Bureau*. Retrieved from http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf

Joseph, L., & Schisler, R. (2009). Should adolescents go back to the basics?: A review of teaching word reading skills to middle and high school students. *Remedial and Special Education, 30*(3), 131-147.

Kober, N. (2010). Improving achievement for the growing Latino population is critical to the nation's future. *Student achievement policy brief #3: Latino students.* Washington, DC: Center on Education Policy.

Lee, S., Olszewski-Kubilius, P., & Peternel, G. (2010). Achievement after participation in a preparatory program for verbally talented students. *Roeper Review, 32*(3), 150-163.

Lopez, E., Gallimore, R., Garnier, H., & Reese, L. (2007). Preschool antecedents of mathematics achievement of Latinos: The influence of family resources, early literacy experiences, and preschool attendance. *Hispanic Journal of Behavioral Sciences, 29*(4), 456-471.

Mucherah, W., & Yoder, A. (2008). Motivation for reading and middle school students' performance on standardized testing in reading. *Reading Psychology*, *29*(3), 214-235.

Norris, P. (2003). *Digital divide: Civic engagement, information poverty, and the internet worldwide*. Cambridge, UK: Cambridge University Press.

Ortiz, R. (2004). Hispanic/Latino fathers and children's literacy development: Examining involvement practices from a sociocultural context. *Journal of Latinos and Education, 3*(3), 165-180.

Partnership for 21st Century Skills. (2011). *Framework for 21st century learning*. Retrieved from http://www.p21.org/index.php?option=com_content&task=view&id=254&Itemid=120.

Pond, L. (1999). *Supporting mathematical development in the early years*. Buckingham, UK: Open University Press.

Poulsen, R., Hastings, P., & Allbritton, D. (2007). Tutoring bilingual students with an automated reading tutor that listens. *Journal of Educational Computing Research, 36*(2), 191-221.

Robinson, J. (2008). Evidence of a differential effect of ability grouping on the reading achievement growth of language-minority Hispanics. *Educational Evaluation and Policy Analysis, 30*(2), 141-180.

Rojas-LeBouef, A., & Slate, J. (2012*). The achievement gap between white and non-white students*. Houston, TX: National Council of Professors of Educational Administration.

Silva, E. (2008). Measuring skills for the 21st century. *Education Sector Reports*, *11*.

Snipes, J., & Horwitz, A. (2008). *Advancing adolescent literacy in urban schools*. Washington, DC: Council of the Great City Schools.

Taningco, M. T. V., & Pachon, H. P. (2008). *Computer Use, Parental Expectations, & Latino Academic Achievement*. Clairmont, CA: Tomas Rivera Policy Institute.

Vassiliou, J. (2011). Improved outcomes with computer-assisted instruction in mathematics and English language skills for Hispanic students in need of remedial education at Miami Dade College, Florida. *Community College Journal of Research and Practice, 35*(3), 191-201

Vogt, W. (2007). *Quantitative research methods for professionals*. Boston, MA: Pearson.

Wamba, N. (Ed.). (2011). *Poverty and literacy*. Florence, KY: Routledge.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., Valdez-Menchaca, M. C., DeBaryshe, B. D., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, *24*(4), 552–559.

Table 1

Descriptive Statistics - Mean Score of Independent and Dependent Variables.

Variable	Mean Score/Correlate
Standardized reading test	50
Race/Ethnicity	n/a
Family composition	n/a
Total family income	$35,000
Student thinks reading is fun	Split on agree/disagree
Geographic region of school	Midwest/South
Access to a computer at home	.88 (0=No/1=Yes)

Table 2

Linear Regression Analysis of variable: "Thinks Reading is Fun" upon DV: Reading Standardized Test Scores.

Hispanic	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
			B	Std. Error	Beta			Lower Bound	Upper Bound
No	1	(Constant)	49.701	.132		377.946	.000	49.443	49.959
		Thinks Reading is Fun	4.272	.187	.220	22.798	.000	3.904	4.639
Yes	1	(Constant)	45.245	.328		137.759	.000	44.601	45.889
		Thinks Reading is Fun	1.448	.463	.076	3.126	.002	.539	2.357

a. Dependent Variable: Reading test standardized score

Table 3

Linear Regression Analysis of variables: "Computer" and "Thinks Reading is Fun" upon DV: Reading Standardized Test Scores.

Hispanic	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
			B	Std. Error	Beta			Lower Bound	Upper Bound
No	1	(Constant)	43.099	.327		131.866	.000	42.458	43.739
		Thinks Reading is Fun	4.340	.183	.224	23.658	.000	3.980	4.700
		Computer	7.264	.328	.209	22.144	.000	6.621	7.907
Yes	1	(Constant)	41.470	.550		75.392	.000	40.391	42.549
		Thinks Reading is Fun	1.531	.457	.080	3.348	.001	.634	2.428
		Computer	4.839	.562	.205	8.607	.000	3.736	5.942

a. Dependent Variable: Reading test standardized score

Table 4

Linear Regression Analysis of variables: "Family Structure," "Thinks Reading is Fun" and "Computer" upon DV: Reading Standardized Test Scores.

Hispanic	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
			B	Std. Error	Beta			Lower Bound	Upper Bound
No	1	(Constant)	41.928	.329		127.394	.000	41.283	42.573
		Thinks Reading is Fun	4.317	.181	.223	23.875	.000	3.962	4.671
		Computer	6.482	.326	.187	19.857	.000	5.842	7.122
		Family Structure	3.215	.186	.163	17.324	.000	2.851	3.579
Yes	1	(Constant)	41.031	.601		68.285	.000	39.853	42.210
		Thinks Reading is Fun	1.537	.457	.080	3.363	.001	.641	2.434
		Computer	4.803	.562	.204	8.542	.000	3.700	5.906
		Family Structure	.833	.460	.043	1.809	.071	-.070	1.736

a. Dependent Variable: Reading test standardized score

Table 5

Linear Regression Analysis of variables: "Urban," "Family Structure," "Thinks Reading is Fun" and "Computer" upon DV: Reading Standardized Test Scores.

Hispanic	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
			B	Std. Error	Beta			Lower Bound	Upper Bound
No	1	(Constant)	42.044	.336		125.205	.000	41.386	42.702
		Thinks Reading is Fun	4.328	.181	.223	23.924	.000	3.973	4.682
		Computer	6.463	.327	.186	19.792	.000	5.823	7.103
		Family Structure	3.197	.186	.162	17.202	.000	2.833	3.561
		Urban	-.358	.206	-.016	-1.735	.083	-.762	.047
Yes	1	(Constant)	41.090	.636		64.622	.000	39.843	42.337
		Thinks Reading is Fun	1.537	.457	.080	3.362	.001	.640	2.433
		Computer	4.800	.562	.204	8.534	.000	3.697	5.903
		Family Structure	.835	.461	.043	1.813	.070	-.068	1.738
		Urban	-.131	.460	-.007	-.284	.776	-1.034	.772

a. Dependent Variable: Reading test standardized score

Table 6

Multi-Category Regression Analysis for DV: Reading Standardized Test Scores

		Model 1		Model 2		Model 3		Model 4	
		B	BETA	B	BETA	B	BETA	B	BETA
Non-Hispanic (No = 0)	Constant/Intercept	49.701		43.099		41.928		42.044	
	IVs								
	Thinks Reading is Fun	4.272	.220	4.340	.224	4.317	.223	4.328	.223
	Family Owns a Computer			7.264	.209	6.482	.187	6.463	.188
	Family Structure					3.215	.163	3.197	.162
	Urban							-.358	-.016
	Adjusted R-squared	.005		.093		.119		.120	
Hispanic (Yes =1)	Constant/Intercept	45.245		41.470		41.031		41.090	
	IVs								
	Thinks Reading is Fun	1.448	.076	1.531	.080	1.537	.080	1.537	.080
	Family Owns a Computer			4.839	.205	4.803	.204	4.800	.204
	Family Structure					.833	.043	.836	.043
	Urban							-.131	-.007
	Adjusted R-squared	.048		.047		.049		.048	

Beyond ORF: Student-Level Predictors of Reading Achievement

Author Info:

Angela I. Canto, Educational Psychology and Learning Systems, Florida State University; Briley E. Proctor, Educational Psychology and Learning Systems, Florida State University; Alicia L. Shafer, Educational Psychology and Learning Systems, Florida State University.

Briley E. Proctor is now at Bainbridge Island School District in Bainbridge Island, WA.
The authors would like to specifically acknowledge and thank the participating volunteer reading coaches, school psychologists, and school district personnel where data were collected. This manuscript is based on data and analyses from the first author's doctoral dissertation.

Correspondence regarding this manuscript should be addressed to Angela Canto, Educational Psychology & Learning Systems, 3210 Stone Building, PO Box 3064453, Tallahassee, FL 32306-4453. Email: acanto@fsu.edu

Abstract

This study explored student-level predictors of reading achievement among third grade regular education students. Predictors included student demographics (sex and socioeconomic status (SES), using free and reduced lunch as proxy for SES), direct observations of reading skills (oral reading fluency (ORF) and word decoding skill (nonsense word fluency/NWF), and academic history (number of prior grade retentions (retentions), Reading/Language Arts grades (reading grade), and attendance rate. Hierarchical linear regression results indicated that ORF and reading grade were statistically significant predictors of high-stakes reading achievement for this sample (model R2=.631). Results replicated previous findings of the predictive value of ORF, above and beyond economic disadvantage and highlighted the influence of low reading grades as an additional key predictor of poor reading achievement, with effect above and beyond that of ORF alone.

Keywords: oral reading fluency, reading, prediction, achievement

Beyond ORF: Student-Level Predictors of Reading Achievement

It is well-known that students' ability to read fluently (accurately, quickly, and with expression) is important for overall academic achievement (e.g., Armbruster, Lehr, & Osborn, 2001; Samuels, 2002). Some degree of automaticity in reading is needed for prompt comprehension of the printed text which helps the reader avoid becoming fixated on pronunciation of isolated words at the expense of understanding the text meaning (Sindelar, Lane, Pullen, & Hudson, 2002; Snow, Burns & Griffin, 1998). Indeed, fluent reading is a known predictor of reading comprehension—the ultimate prize or purpose for reading—with correlations between reading fluency and comprehension ranging between .70 and .90 (Baker, Gersten, & Grossen, 2002). Research consistently indicates that Oral Reading Fluency (ORF)—reading connected text aloud—is a critical indicator of general reading skill (Fuchs, 1995). When teachers use ORF data to establish individual student achievement goals, monitor the effects of instructional programs, and adjust interventions accordingly, student achievement improves (Connor, Morrison, & Petrella, 2004; Shinn, 1995; Shinn, Shinn, Hamilton, & Clarke, 2002; Stecker, & Fuchs, 2000).

ORF measures generally demonstrate strong overall technical adequacy (i.e., reliability and validity) (e.g., Deno, 1985, 1989; Fuchs, 1995; Fuchs, Fuchs, & Maxwell, 1988; Good & Jefferson, 1998; Hosp & Fuchs, 2005; Marston, 1989). As cited in these studies and Marston (1989), reliability measures are generally high with most estimates of test-retest reliability (ranging from .82 to .97) and parallel forms reliability (ranging from .84 to .96) being above .90. Inter-rater reliability estimates for ORF procedures have been achieved at .99 (Tindal, Marston, & Deno, 1983 as cited in Marston, 1989). In validity studies, researchers have concluded that ORF assessment procedures appear to result in data possessing adequate to strong validity overall (Fuchs et al., 1988; Marston, 1989). Additionally data obtained through ORF procedures appear to possess moderate to strong concurrent and discriminant validity with other measures of reading skill including oral passage reading, question-answering tests, recall of text procedures, cloze procedures of reading comprehension (i.e., missing word completion measure), and broader measures of reading comprehension (Fuchs et al., 1988).

Student ORF scores have been used to predict reading achievement on many state adopted criterion-referenced tests of achievement (e.g., Buck & Torgesen, 2003; Hixson & McGlinchey, 2004; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shapiro, Keller, Lutz, Santoro & Hintze, 2006; Silberglitt, Burns, Madyun, & Lail, 2006; Wanzek, Roberts, Linan-Thompson, Vaughn, Woodruff, & Murray, 2010) as well as nationally norm-referenced tests of achievement (Hixson & McGlinchey, 2004; Klein & Jimerson, 2005; Roehrig, et al., 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Wanzek et al., 2010). The proportion of variance explained by ORF in these studies tends to fall between 36% (e.g., Wanzek et al., 2010) and 64% (e.g., Hixson & McGlinchey, 2004), depending on the study and the predictor variables included in the model. Notably, Kranzler, Brownell, and Miller (1998) reported that ORF is not simply a proxy for underlying cognitive processes including cognitive ability, processing speed, and efficiency but rather contributes unique

variance to the prediction of reading achievement.

One limitation in using ORF, however, is that studies of ORF predictive validity have had mixed results among some ethnic minority subgroups and students of low socioeconomic status (e.g., Buck & Torgesen, 2003; Crowe, Connor, & Petscher, 2009; Hintze, Callahan, Matthews & Williams, 2002; Hixson & McGlinchey, 2004; Hosp, Hosp, & Dole, 2011; Klein & Jimerson, 2005; Kranzler, Miller, & Jordan, 1999). Recently, Hosp, Hosp and Dole (2011) called for additional research noting that while the predictive validity of ORF was generally quite good, it "may not demonstrate consistent levels of predictive validity when focusing on different subgroups" (p. 125). Hosp and colleagues (2011) suggest that the source of this "predictive bias" is difficult to pinpoint. They offered several possible explanations, including the possibility that differences were the result of *a priori* decisions regarding variables included in the prediction models. In sum, ORF research suggests that it is a good overall predictor of reading achievement but that caution may be warranted when interpreting the predictive validity for specific subgroups. The research on predictive validity of ORF may need additional studies to determine the overall pattern (Hosp et al., 2011).

Efforts to improve the prediction of reading achievement by the inclusion of other student-level variables have been rare. The study by Hosp and colleagues (2011), for example, appears to be the only published report examining the relationship between word decoding skill in third grade and third grade high-stakes reading achievement. This is somewhat surprising because it has long been argued that, in addition to oral reading fluency, decoding is also a requisite skill requisite for success on high-stakes measures of reading achievement (Armbruster et al., 2001; Marston, 1989). In fact, text passages on year-end reading achievement tests often include higher-level decodable words (Hiebert, 2002) and decoding ability has been found to be a reliable indicator of persistent reading difficulties (Burke, Hagan-Burke, Kwok, & Parker, 2009). Thus, a measure of decoding may have utility for enhancing prediction of high-stakes reading achievement, but is yet unknown.

In addition to ORF and decoding, researchers are encouraged to explore additional variables that may enhance prediction of student reading achievement. Bishop and League (2006) highlight the importance of using a multivariate screening model of reading achievement. At this time, however, we know little about the impact of other student-level variables on reading achievement. Other variables such as students' reading grades, attendance rate, and prior grade retentions may also explain a significant portion of variance in high-stakes reading achievement scores above and beyond that of ORF. For example, research has shown only rare support for mean differences between sexes on ORF and norm-referenced measures (second grade spring differences between sexes on ORF; Klein & Jimerson, 2005), yet, sex differences have been documented on student grades (Burts, Hart, Charlesworth, & DeWolf, 1993) and grade retention (Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997; McCoy & Reynolds, 1999). Additionally, variables such as grades and prior grade retentions seem to have intuitive relationships with reading achievement overall; yet, whether the effects of those variables explain additional significant variance over ORF is unknown.

In summary, the purpose of the present study was threefold. First, we were interested in replicating earlier studies on the prediction of high-stakes reading achievement among third grade students using ORF while controlling for student demographics (economic disadvantage and sex). Students' free and reduced lunch status was used as a proxy for SES. It was hypothesized that our findings would be consistent with those reported in earlier investigations on the predictive utility of ORF, controlling for student demographics. Secondly, we wanted to test whether the inclusion of a measure of student decoding would help to improve the prediction model, given that the literature suggests that decoding may still be a factor on achievement on year-end high stakes reading tests. Thirdly, we wanted to explore whether prediction of high-stakes reading achievement among third grade students could be enhanced by the inclusion of additional student-level variables known to be implicated in overall school achievement. Thus, we included in the model data on the student's number of prior grade retentions, attendance rate, and reading grade. These final three variables are data that are readily available to teachers and do not require time or resources for additional direct measurement of student skill. It was hypothesized that the inclusion of these additional student-level variables would increase the proportion of explained variance in the prediction of reading achievement scores.

Methods

Participants

Third grade students (n = 145) in a large southeastern school district participated in this investigation. This large metropolitan school district subdivided their schools into five district regions. There is variability in student demographics across these district regions, especially with regard to ethnic diversity and SES (using free and reduced lunch status as a proxy family income indicator). Four elementary schools from each of the five district regions were recruited in order to capitalize on the naturally occurring ethnic and SES variability in the different geographical locations. Both high and low-performing schools with respect to students' scores on the previous year's statewide high-stakes assessment were intentionally selected to ensure variability in achievement scores. Of 20 schools invited, 12 principals subsequently agreed to participate in the present study. Each regular education third grade teacher within the participating schools was then individually invited and all subsequently agreed to participate. Students were eligible for participation if they were enrolled in the participating teacher's classroom as a regular education student.

The required sample size to detect a large effect (Cohen's d = 0.8) was calculated based on a two-tailed linear multiple regression (random model) with a confidence level of .95 and a statistical power of .80 and 8 predictor variables, indicating that the researchers needed to obtain at least 102 participants (Faul, Erdfelder, Buchner, & Lang, 2009). Acknowledging the potential for a low return rate of consent forms, 32 regular education third grade students were randomly selected from each of the 12 schools using a random numbers chart and a total of 384 consent packets as approved by our university's institutional review board were sent home with students in their backpacks. Of the 384 informed consent packets distributed, 192 consent forms (or 50.3 %) were returned. Of those received, 186 parents/legal guardians consented (96.9% of consent forms returned), 6

parents/legal guardians declined participation (3.1%). Five participants were no longer enrolled in the participating school at the conclusion of the study (2.7% attrition). The demographic composition of the final sample is summarized in Table 1. Socio-economic status (SES) was characterized in this study via a dichotomous variable: *economically disadvantaged* (i.e., students receiving either free or reduced lunch price benefits) and *non-economically disadvantaged* (i.e., students that did not apply or were ineligible for free or reduced lunch price benefits).

Only students for whom reading skill performance data (ORF & decoding/Nonsense Word Fluency (NWF)) were available were retained in the final analysis, resulting in 145 cases for analysis. It was determined that the loss in sample size and concomitant loss in power in eliminating cases with missing data was preferable over imputing those values. Thus, the multiple regression results are based on data from 145 participants.

Instruments

Instruments used in the present study included a year-end high-stakes measure of reading progress for grade 3 (*Florida Comprehensive Assessment Test; FCAT*), the ORF (oral reading fluency) and NWF (nonsense word fluency) subtests from *Dynamic Indicators of Basic Early Literacy Skills* (*DIBELS*) assessment system (Good & Kaminski, 2002), and a brief survey administered to each participant's teacher to obtain the participants' third quarter reading grade. ORF and NWF subtests were used in unaltered form from the *DIBELS* assessment system (Good & Kaminski, 2002) and administered and scored following the standardized administration and scoring procedures provided for the instrument. Technical adequacy of ORF is reported above; information regarding NWF and *FCAT* is described below. The remaining data (e.g., demographics, attendance) were obtained via query to the district's student database records.

NWF is a decoding task whereby the student reads aloud a series of vowel-consonant or consonant-vowel-consonant nonsense words. This subtest assesses the student's ability to blend phonemes, requiring both knowledge of letter-sound correspondences and articulation skill. First grade January NWF scores appear to possess strong predictive validity for end-of-first-grade ORF scores (.82) (Good & Kaminski, 2002). Predictive validity appears weaker for end-of-second-grade ORF scores (.60) and for the Woodcock-Johnson Psycho-Educational Battery (Woodcock, McGrew, & Mather, 2001) Total Reading Cluster score (.66). The instrument's authors did not intend for the NWF subtest to be administered to third grade students and, therefore, there is currently no data to examine the reliability, validity, and predictive utility for this grade level. Nonetheless, as discussed, we were specifically interested in including a measure of decoding given that it is a requisite skill for overall reading achievement of new words, especially for struggling readers in third grade. For this study, the second grade benchmark NWF probes were used intact with no modifications.

Student scores from the *FCAT* Reading subtest were used as a general measure of reading achievement consisting of 50 to 55 multiple choice questions at the time this study was conducted. Students were provided informational (subject-matter centered) or literary (fiction, nonfiction, poetry, or drama) text passages and asked to answer questions to assess students' ability

to construct meaning from the texts. Scores on the *FCAT* are reported in terms of scaled scores (range 100-500) and achievement level (range 1-5) (Florida Department of Education (FDOE), 2001, 2004). The parallel forms reliability for the *FCAT* was above .90 for grades 4, 5, 8, and 10 (FDOE, 2001) and correlations between the *FCAT* and SAT-9 two measures ranged from .70 to .81 (FDOE, 2001). The mean Reading *FCAT* for third grade standard curriculum students (non-ESE students) was 317.22 (sd = 56.97) for the year in which this study was conducted. Reliability as measured by Cronbach's alpha was strong at .89 for this administration of the *FCAT*.

Procedures

ORF and NWF subtests were administered within a two-week interval in early December, approximately 14-16 weeks prior to the springtime high-stakes assessment of reading achievement. Volunteer school psychologists and school-based reading coaches administered the subtests, all of whom had received a minimum of six hours of formal in-service training in the administration and scoring of the selected *DIBELS* subtests. Each participant was read a scripted assent form prior to administration.

Twenty percent (n = 36) of the protocols from both subtests were randomly selected for reliability checks by the lead author. Results of the reliability checks are as follows: NWF = .72; ORF = 1.00. Errors were noted in the scoring of NWF, including addition errors, neglect of reporting the maximum correct number of phonemes per line, and omission of completion time if under 1 minute. The lead author re-scored each NWF protocol and NWF protocols that did not note completion (8.8%; n = 16 of 181 students tested) were deemed spoiled and eliminated from analysis.

Sex, SES, attendance rate, and number of prior grade retentions were retrieved from the school district's database. Student attendance rate was obtained by dividing the number of days the student was enrolled by the number of days the student was present for the academic year. The sample median attendance rate was .97 (IQR = .039). With regard to grade retention, of the 145 students used in the regression analysis, 37 students (25.5%) had been retained at least once. Of those retained, 10 (6.9%) were retained in Kindergarten, 12 (8.3%) in first grade, 8 (5.5%) in second grade, and 33 (22.8%) in third grade. Twenty-seven of those students had been retained once, 10 students retained twice. An additional 16 students were retained at the conclusion of the study (14 of whom failed the *FCAT*).

Teachers were provided a questionnaire on which to report each participant's third quarter reading grade with self-addressed stamped envelopes provided for return. Of those distributed, 28.2% of the questionnaires were not returned. The school district database only retained the final reading grade for the academic year, deleting the 9-week quarter grades from the database. Therefore, in cases where the third quarter grade was unavailable, the final reading grade was used.

The purpose of this study was: 1) to replicate earlier studies using ORF to predict reading achievement among third grade students, while controlling for student demographics (economic disadvantage and sex); 2) to test whether the inclusion of a measure of student decoding would help to improve the prediction of reading achievement; and 3) to test whether the inclusion of additional student-level

variables known to be implicated in overall school achievement—student's number of prior grade retentions, attendance rate, and reading grade—improve the prediction model. While there are several possible avenues of analysis one could use to explore these questions, hierarchical regression was utilized to better understand the individual and additive effects of each predictor variable or variable set.

When interpreting the results from hierarchical regression analyses, the order of entry of variables into the model should be based on sound empirical or theoretical reasoning (Keith, 2006). While several alternatives exist, the following order was used to address the stated purposes of this study. SES and sex were entered in the first block as control variables to control for the effects of these demographics on achievement. ORF was then entered second into the model to determine its effect on reading achievement when controlling for the aforementioned student demographics (replication of prior studies). NWF was entered third in the model to test the added predictive value of decoding on the reading achievement test, above and beyond that of ORF. The remaining student level variables (retentions, attendance rate, and reading grade) were then entered into the fourth and final block to explore the whether the inclusion of these additional student-level variables would increase the proportion of explained variance in the prediction of reading achievement scores above and beyond demographics, ORF, and decoding skill.

Results

The inter-correlation matrix of predictors is provided in Table 2 with associated tests of significance of the relationships between variables using $\alpha = .01$. Significant correlations were found between the *FCAT* reading measure and ORF, NWF, SES, number of prior grade retentions (retentions), and reading grades. Of interest, the significant negative correlation between the reading *FCAT* score and retentions indicated that students who were retained one or more times performed significantly poorer on the outcome reading measure. With regard to student demographics, SES was significantly correlated with ORF, NWF, retentions, and reading grade indicating that students with economic disadvantage were significantly more likely to perform worse on ORF and NWF measures, had been retained at least once, and had poorer reading grades than the group of students that were categorized as not economically disadvantaged. Sex was not significantly correlated with any other variables included in the model. ORF was significantly positively correlated with NWF and reading grades and significantly negatively correlated with retentions. Similarly, NWF was significantly positively correlated with reading grades.

Hierarchical Regression Results

A case analysis was conducted to evaluate the presence of potential outliers exerting excessive influence on the regression results. One outlier was identified; however, a subsequent sensitivity study revealed that the outlier was not exerting excessive influence on the model R^2 (change in R^2 = .011). Thus, the observation was retained and the reported results reflect the inclusion of all participant data (n=145). The model was run with all variables, retaining the studentized model residuals. A visual inspection of the scatter plot of the studentized model residuals versus predicted Y values revealed no indications of any violations of correct fit of a linear model, constant variance, or normality assumptions

required for the legitimacy of the regression results.

Hierarchical linear regression results are provided in Table 3. In summary, the addition of the demographic controls into the first block revealed that only SES was significantly predictive of Reading *FCAT* scores, $\Delta R^2 = .192$, $F(2, 142) = 16.92$, $p < .001$. The addition of ORF into the second block confirmed the significance of the relationship between ORF on Reading *FCAT* scores even when controlling for student demographics (SES and sex), $\Delta R^2 = .334$, $F(1, 141) = 52.26$, $p < .001$. The effect of SES in the second block remained statistically significant ($p < .001$). The addition of the third block revealed no significant effect of NWF, above and beyond ORF and student demographics, $\Delta R^2 = .003$, $F(1, 140) = 39.39$, $p < .35$, while the significance of SES and ORF remain unchanged. In the fourth and final block, the addition of the remaining student-level variables (retentions, attendance rate, and reading grades) significantly increased the explained variance in *FCAT* Reading, $\Delta R^2 = .102$, $F(3, 137) = 33.47$, $p < .001$, with reading grade offering the only significant unique contribution in this step ($p < .001$). In the final model SES was not significant (albeit marginally) while the significant effect of ORF observed in earlier blocks remained significant. Overall, results indicated that a multivariate model of student level predictors of reading achievement was robust and an improvement over a model that included ORF and student demographics in isolation. The final model with all predictor variables resulted in a model $R^2 = 0.631$ (Adjusted $R^2 = 0.612$), a fairly large coefficient of determination, indicating that approximately 63% of the total variability of reading achievement scores could be explained using this model.

Discussion

The relevance of this paper rests in the use of regression to identify additional student-level factors, including word decoding skill, academic history, and demographics that may contribute to success on a comprehensive statewide third grade reading achievement test above and beyond ORF. The authors hypothesized that the inclusion of additional student-level variables would improve the overall model, increasing the proportion of variance in reading achievement. This hypothesis was supported. In the final model, the effects of ORF and reading grades were significant in predicting year-end reading achievement, significant in spite of including student economic disadvantage and prior grade retentions into the prediction model—two factors often implicated in poor reading achievement. Thus, these results are encouraging given that it points to factors that can be addressed with instruction and may help ameliorate learning or performance deficits associated with disadvantage and grade retention issues.

As such, ORF continues to be an important factor in predicting reading achievement and continued focus on students' ability to read fluently appears warranted. Armbruster, Lehr, and Osborn (2001), report that the use of frequent oral reading monitored by a teacher or parent (coined "repeated readings") is an effective activity for improving reading fluency and overall reading achievement. Results herein support continued use of interventions that would target reading fluency, perhaps using locally-derived or state-derived benchmarks rather than national benchmarks for ORF to predict year-end assessment success may be more useful (Brown, 2008). In their 2001 study, Crawford, Tindal and Stieber found that "a reading rate of 119 words per minute

virtually ensured that a student passed the [Oregon] statewide reading test" (p. 319). This equated to 94% of their sample. In this study, 93% of students that read 113 correct words per minute on ORF subsequently achieved a passing score on the state's year-end reading assessment. It is yet unclear how often individual teachers are establishing and using local benchmarks.

A measure of decoding was specifically added to this study with the hypothesis that decoding remains an important skill for success on high stakes year end reading assessments, which typically include higher-level decodable words. Of note, NWF is not typically administered in third grade and benchmarks are unavailable for this period (Good & Kaminski, 2002). Thus, the inclusion of NWF in this study with third graders was exploratory; however, the hypothesis for its importance was not supported by the data. A post-hoc analysis revealed that if entered first, NWF was a significant predictor (as would be expected from the correlation matrix), but its effect was negated as soon as ORF entered the model. The insignificance of NWF could be contributed to the high correlation between ORF and NWF such that NWF did not add any unique contribution. Nonetheless, NWF simply did not appear to be an important factor independent of the effect of ORF in third grade.

In contrast, the finding that reading grades did uniquely contribute to the prediction of reading achievement above and beyond ORF, and in the context of all of the other variables in the model was unexpected. The predictive utility of teachers' assigned reading grades has not been widely discussed in the literature on predicting reading achievement. It is conceivable that students' reading fluency skill in general contributed, at least in part, to the letter grades assigned to students for reading/language arts; however, by putting ORF in the hierarchical regression analysis first we were then able to explore how grades *added* to that predictive power. It is indeed likely that participating teachers at different schools (or even within a school) may use different criteria to determine a student's reading grade (e.g., may include data on students' participation, work completion, vocabulary, and spelling tests). Nonetheless, ORF alone was not as predictive of reading achievement as was a model that included reading grades.

A *post-hoc* analysis revealed that approximately 4% of the participants in this study who earned a third-quarter reading grade of an A failed the Reading *FCAT*. Moreover, approximately 8% of those who earned a B failed, 44% who earned a C failed, 67% who earned a D failed, and 78% who earned an F failed the state's year end assessment of reading achievement. Perhaps in teachers' constructions of reading grades the teachers are picking up on something above that of reading fluency which is contributing to overall reading achievement. Previous research regarding accuracy of teachers' assessments of reading skill indicates that teachers are good judges of a variety of reading skills. Feinberg and Shapiro (2003) noted that students' assessed oral reading fluency skill was highly correlated with teachers' predictions of oral reading fluency rate (r = .70). Additionally, Bates and Nettelbeck (2001) examined the accuracy of teacher judgments in reading achievement among students with and without classroom behavioral problems. Results of this study indicated that teachers remained accurate judges of reading accuracy (r = .77) and reading comprehension (r = .62), despite the presence of classroom behavior problems. While students with behavior problems

tended to perform more poorly on the reading measures, the teachers did not underestimate reading skill (Bates & Nettelbeck, 2001).

In sum, results of this study: 1) support previous findings of the predictive value of ORF, even when controlling for student demographics; 2) do not support the use of a decoding measure to improve the prediction of reading achievement; and 3) highlight the unique influence reading grades on the prediction of reading achievement. From a cost-benefit analysis perspective, over-identifying students as needing additional intervention may be preferable than under-identifying under-achieving students (Roehrig et al., 2008). Glover and Albers (2007) discuss both pros and cons to over and under-identification (e.g., increased burden on programming resources), but agree that under-identification is a greater risk when the consequences are more high-stakes, such as in year-end achievement testing.

Limitations

Limitations in this study were rooted primarily in the lack of consenting participants, missing data, and subtest administration adherence. Although the return rate of consent packets sent home to parents was consistent with average return rates for mailed or sent-home documentation, the final sample may still represent systematic bias toward families who are possibly more involved in their children's education or more conscientious in completing requested documentation. While results may not be universally generalizable to all third grade students, the random sampling method was a strength in this study.

Missing data also posed some difficulty for this study; teacher survey return rates were not as high as expected, resulting in missing third quarter reading grades for several students. To compensate, the final reading grades were used to replace missing values as described above. Additionally, of 181 consenting participants, ORF and NWF data were collected for 145 participants with absenteeism as the most common cause of missing assessment data. While a variety of strategies and statistical techniques are available to researchers, each with pros and cons (see Baraldi and Enders, 2010), we elected to retain the 145 cases that had both ORF and NWF data, accepting the minimal loss in statistical power. Nonetheless, using the 145-case subset still could have created a biased subset of the original 181 cases and is offered as a limitation of this study.

Subtest administration error for NWF was mildly problematic in the present study (8.8% of protocols with errors). The NWF subtest is used less frequently than ORF in this district and is reported by some testers as more difficult to administer and score given that exact pronunciation of individual phonemes is required for score credit. Perhaps including protocol for inter-observer reliability checks during administration would have been helpful in pinpointing the specific source of the problems associated with that measure.

Lastly, with 24 of the third grade students in the sample repeating third grade, it is possible that these students had seen the oral reading fluency passages used in this sample at some point prior to this study, potentially affecting the results. However, this is unlikely given that these students (who were already not achieving well academically) were able to decode, comprehend, and/or recall the passages in any great detail that would substantially alter the ORF scores for

those passages. Additionally, latter passages of the benchmark assessment probes were used in this study that at that time were not being used by the schools for progress monitoring.

Future Directions

While this study targeted regular education third grade students, it may be of interest to replicate the present study with a larger, more diverse sample, increasing generalizability of results and allowing for additional comparisons within and across grades with the identified predictors. With a larger, more diverse sample, one might also explore the prediction model for specific subgroups. It may be useful to analyze current reading risk models using ORF benchmarks to delineate cutoffs that appropriately identify students at risk for failure across groups, with due caution in interpretation of any group differences. It is plausible that the significant predictor variables for students who are English language learners (Wiley & Deno, 2005; Yeo, 2010) or students with Specific Learning Disabilities may differ than those found to be significant in the present study with regular education students.

The present study was focused on Reading *FCAT* achievement. Previous research by Buck and Torgesen (2003) examined the correlation between ORF and Math *FCAT* achievement as well and found a significant positive correlation between the two (r = .54, $p < .001$). Similarly, the predictor variables in the present study could be applied to predict math achievement on year end measures of achievement. In lieu of ORF, using silent curriculum-based measures of reading such as maze measures may also prove useful in predicting state assessments of math achievement (Jiban & Deno, 2007). Lastly, it would be interesting to further dissect reading grades such that we can better understand how teacher evaluations map onto reading skills that are important for grade level assessments of reading achievement. This area appears to be ripe for further research.

References

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read*. Washington, DC: Partnership for Reading.

Baker, S., Gersten, R., & Grossen, B. (2002). Interventions for students with reading comprehension problems. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 731-754). Bethesda, MD: National Association of School Psychologists.

Baraldi, A. N. & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48,* 5-37. doi:10.1016/j.jsp.2009.10.001

Bates, C. & Nettelbeck, T. (2001). Primary school teachers' judgments of reading achievement. *Educational Psychology, 21,* 177-187. doi:10.1080/01443410020043878

Bishop, A. G. & League M. B. (2006). Identifying a multivariate screening model to predict reading difficulties at the onset of Kindergarten: A longitudinal analysis. *Learning Disability Quarterly, 29,* 235-252.

Brown, S. E. (2008). Comparing methodologies used to predict outcomes on high stakes tests with curriculum-based measurement. *Dissertation Abstracts International Section A: Humanities and Social Sciences.*

Buck, J. & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report No. 1). Tallahassee, FL: Florida Center for Reading Research and Florida State University. Retrieved from http://www.fcrr.org/technicalreports/TechnicalReport1.pdf

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of Kindergarten to second grade. *The Journal of Special Education, 42,* 209-226. doi:10.1177/0022466907313347

Burts, D. C., Hart, C. H., Charlesworth, R., & DeWolf, D. M. (1993). Developmental appropriateness of kindergarten programs and academic outcomes in first grade. *Journal of Research in Childhood Education, 8*, (1) 23-31.

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child by instruction interactions. *Journal of Educational Psychology, 96*(4), 682-698. doi:10.1037/0022-0663.96.4.682

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323. doi:10.1207/S15326977EA0704_04

Crowe, E. C., Connor, C. M. & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47,* 187-214. doi:10.1016/j.jsp.2009.02.002

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based* measurement: Assessing special children (pp. 1-17). NY: The Guilford Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Feinberg, A. B. & Shapiro E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18,* 52-65. doi: 10.1521/scpq.18.1.52.20876

Florida Department of Education (2001). *FCAT briefing book*. Retrieved August 13, 2003, from http://www.firn.edu/doe/sas/*FCAT*/pdf/*FCAT*_brief.pdf.

Florida Department of Education (2004). *Assessment and accountability briefing book*. Retrieved January 16, 2006, from http://www.firn.edu/doe/sas/*FCAT*/pdf/*FCAT*aabb.pdf

Fuchs, L. S. (1995). Best practices in defining student goals and outcomes. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 539-546). Washington, DC: The National Association of School Psychologists.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20-28. doi: 10.1177/074193258800900206

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117-135. doi:10.1016/j.jsp.2006.05.005

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp.61-88). NY: The Guilford Press.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 6th Edition.* Eugene, OR: Institute for Development of Educational Achievement. (Available at http://*DIBELS*.uoregon.edu/)

Hiebert, E. H. (2002). Standards, assessments, and text difficulty. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.) (pp. 337- 369). Newark, DE: International Reading Association.

Hintze, J. M., Callahan, J. E., III, Matthews, W. J., & Williams, S. A. S. (2002). Oral reading fluency and prediction of reading comprehension in African-American and Caucasian elementary school children. *School Psychology Review, 31,* 540-553.

Hixson, M. D. & McGlinchey, M. T. (2004). The relationship between race, income, and oral reading fluency and performance on two reading comprehension measures. *Journal of Psychoeducational Assessment, 22,* 351-364. doi: 10.1177/073428290402200405

Hosp, M. K. & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word-reading, and comprehension: do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40*(1), 108-131.

Jiban, C. L. & Deno., S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance. *Assessment for Effective Intervention, 32*(2), 78-89. doi: 10.1177/15345084070320020501

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30,* 420-437.

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology, 35*(1), 3-25. doi:10.1016/S0022-4405(96)00033-7

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Allyn & Bacon.

Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly, 20,*(1), 23-50. doi:10.1521/scpq.20.1.23.64196

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36,* 399-415. doi:10.1016/S0022-4405(98)00018-1

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly 14*(3), 327-342. doi:10.1037/h0089012

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). NY: The Guilford Press.

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*(3), 273-298.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the *DIBELS Oral* Reading Fluency measure on predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343-366. doi:10.1016/j.jsp.2007.06.006

Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.) (pp. 166-183). Newark, DE: International Reading Association.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*(5), 429-448. doi:10.1086/51862

Shapiro, E. S., Keller, M. A., Lutz, L. G., Santoro, L. E., Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35. doi: 10.1177/0734282905285237

Shinn, M. R. (1995). Best practices in curriculum-based measurement and its use in a problem solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 547-567). Washington, DC: The National Association of School Psychologists.

Shinn, M. R., Shinn, M. M., Hamilton, C., & Clarke, B. (2002). Using curriculum-based measurement in general education classrooms to promote reading success. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 113-142). Bethesda, MD: National Association of School Psychologists.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: a longitudinal comparison of grade levels. *Psychology in the Schools, 43,* 527-536. doi:10.1002/pits.20175

Sindelar, P. T., Lane, H. B., Pullen, P. C., & Hudson, R. F. (2002). Remedial Interventions for Students with Reading Decoding Problems. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 703-729). Bethesda, MD: National Association of School Psychologists.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15,* 128-134.

Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruff, A. L., & Murray, C. S. (2010). Differences in the relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessment for Effective Instruction, 35,* (2) 67-77. doi: 10.1177/1534508409339917

Wiley H. I. & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* (4) 207-214. doi:10.1177/07419325050260040301

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.

Yeo, S. (2010). Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education, 31,* 412-422. doi:10.1177/0741932508327463

Table 1

Descriptive Statistics

Variable	*n*	*%*
Sex		
Male	55	37.9
Female	90	62.1
Ethnicity		
Caucasian	51	35.2
African-American	87	60.0
Hispanic	3	2.1
Asian	2	1.4
Other	2	1.3
Socio-economic status (SES)		
Disadvantaged	88	60.7
Non-disadvantaged	57	39.3
Number of Retentions		
0	108	74.5
1	27	18.6
2	10	6.9
Reading Grade		
A	23	15.9
B	51	35.2
C	41	28.3
D	23	15.9
F	7	4.8

Note. Reading grade statistics incorporate replaced values for missing third quarter data.

Table 2
Descriptive Statistics and Inter-correlation Matrix

	FCAT	SES	Sex	ORF	NWF	Retentions	Attendance	Reading Grade
FCAT	—							
SES	-.43*	—						
Sex	-.07	.02	—					
ORF	.69*	-.32*	-.07	—				
NWF	.50*	-.30*	.10	.63*	—			
Retentions	-.36*	.25*	.17	-.29*	-.19	—		
Attendance	.04	-.07	-.03	.00	.06	.01	—	
Reading Grade	.65*	-.36*	-.05	.49*	.42*	-.28*	.05	—
M	307.82	.61	.38	102.61	81.55	.32	.96	2.41
SD	52.43	.49	.49	33.54	41.98	.60	.04	1.08
n	145	145	145	145	145	145	145	145
Range	134 - 446	—	—	0 - 180	6 - 232	0 - 2	.68 – 1.0	0 - 4

Note. Correlations for FCAT, ORF, NWF, Retentions, Reading grade, and attendance are the Pearson product-moment correlation. Point estimates for the dichotomous variables of Sex and SES are the contrast of means between the two groups. Variables are coded as follows: Sex: female=0 and male=1; SES: non-disadvantaged=0 and disadvantaged=1; Reading Grade: F=0, D=1, C=2, B=3, A=4.
* $p < .01$.

Table 3
Hierarchical Linear Regression Results Predicting Reading Achievement

	Model 1			Model 2			Model 3			Model 4		
Variable	B	β	*p*	B	β	*p*	B	β	*p*	B	β	*p*
Step 1:												
SES	-46.34	-.43	<.001	-25.71	-.24	<.001	-24.85	-.23	<.001	-15.22	-.14	.014
Sex	-6.66	-.06	.414	-2.49	-.02	.691	-3.60	-.03	.573	-.62	-.01	.915
Step 2:												
ORF				.96	.61	<.001	.89	.57	<.001	.67	.43	<.001
Step 3:												
NWF							.09	.07	.349	.02	.02	.784
Step 4:												
Retentions										-8.21	-.09	.098
Attendance										5.81	.005	.924
Reading Grade										17.10	.35	<.001
R^2	.19			.53			.53			.63		
Change in R^2				.33			.00			.10		
F	16.92***			52.26***			39.39***			33.47***		
Change in *F*				99.48***			.88			12.57***		

Note. The ΔR^2 result is the increase in R^2 due to adding each block incrementally, given the other predictor variables. Variables are coded as follows: Sex: female=0 and male=1; SES: non-disadvantaged=0 and disadvantaged=1; Reading Grade: F=0, D=1, C=2, B=3, A=4.
* $p < .05$. ** $p < .01$. *** $p < .001$

Exemplary Teachers of English Language Learners: A Knowledge Base

Author Info:

Courtney Clayton is an Assistant Professor in Curriculum and Instruction at the University of Mary Washington, Fredericksburg, VA.

Correspondence regarding this manuscript should be addressed to Courtney Clayton, Department of Curriculum & Instruction, Trinkle Hall 217, University of Mary Washington, 1301 College Avenue, Fredericksburg, VA 22401. Ph: 540-654-1360. Email: cclayton@umw.edu

Abstract

The number of English language learners (ELLs) in schools continues to rise. However, statistics reveal that the majority of classroom teachers have no training in working with ELLs (NCES, 2011). Because of this, it is critical to understand how teachers can be successfully prepared to teach ELLs. Through in-depth inquiry, this study explored what kinds of knowledge four exemplary teachers of ELLs had, how they gained that knowledge, and how the knowledge was influenced by the context in which they taught– bilingual or monolingual. A knowledge base for what it takes to be a successful teacher of ELLs is presented.

The number of English language learners (ELLs) in public schools in the United States continues to rise. Between 1980 and 2009, this number rose from 4.7 million to 11.2 million, or from ten to 21 percent (NCES, 2011). However, national statistics reveal that the majority of classroom teachers have little to no training in working with ELLs (NCES, 2011). Therefore, though teachers will increasingly be faced with teaching ELLs in their classrooms, most will be unprepared to deliver effective instruction to these students. In addition, many states have done away with bilingual programs aimed at instructing ELLs in both English and their native languages in favor of English-only models (Proposition 227, California Secretary of State, 1998; Proposition 203, Arizona Secretary of State, 2000; General Laws of Massachusetts, 2002), which has meant that more ELLs are now being fully incorporated into English-only programs leaving teachers feeling overwhelmed by how to best educate these students (Palmer & Garcia, 2000; Stritikus & Garcia, 2000).

In light of this, it is critical to understand how teachers can be successfully prepared to teach ELLs. While there is research on effective practices for ELLs (August & Shannahan, 2006; Coyne, Kame'enui & Carnine, 2011; Echevarria, Short & Vogt, 2007; Tellez & Waxman, 2006), little research has been done examining the knowledge base of successful teachers of ELLs which includes what they actually do in their classrooms, how they learned to do what they do, and what has influenced and facilitated their success. In addition, there has been limited research looking at the role context – bilingual or monolingual - plays in teachers working with ELLs (Garcia, 1991; Tikunoff, 1983).

Through in-depth inquiry with exemplary teachers of ELLs across bilingual and monolingual contexts, this study explored what kinds of knowledge these teachers had in order to successfully work with ELLs, how they gained that knowledge, and how the knowledge was influenced by the context in which teachers worked – bilingual or monolingual. A knowledge base for what it takes to be a successful teacher of ELLs specifically, is presented.

Theoretical Framework

The current study was grounded in the framework developed by Bransford, Darling-Hammond and LePage (2005) for understanding teacher knowledge. Their framework had three intersecting components: 1) knowledge of learners and their development within social contexts, 2) knowledge of subject matter and curriculum goals in light of the social purposes of education, and 3) knowledge of teaching in light of the content and learners to be taught, as informed by assessment and supported by classroom environments. With regard to the knowledge of teaching component, one area that was focused upon was the teaching of diverse learners. Part of the process in learning how to teach diverse learners, they argued, was understanding and reaching out to children who have a wide range of life experiences, behaviors and beliefs. It means understanding that "individuals' world views are not universal but are greatly influenced by their gender, race, ethnicity and social-class background" (p.36). Under this category of "diverse learners" are ELLs; however, ELLs have particular needs in terms of language development beyond what other diverse learners may need, and this was not specifically addressed in the Bransford model. Therefore, in the current study, the knowledge base necessary for teaching ELLs was studied in depth. As Lucas and Grinberg (2008) argue, "It is time

that we stop subsuming the preparation of classroom teachers to teach English language learners within more general considerations of the preparation of teachers for diverse populations" (p. 606). In addition, this study addressed the role of context, as described by Bransford et. al. (2005), in examining the knowledge base of successful teachers of ELLs.

Review of the Literature

Teacher Knowledge

The role of teacher knowledge and how it has been defined are important to the current study since it focused on what exemplary teachers of ELLs know about teaching these students, and how they developed this knowledge. Teacher knowledge has been defined in multiple ways by scholars (e.g. Clandinin & Connelly, 1995; Cochran-Smith & Lytle, 1999; Shulman, 1987). Grimmett & MacKinnon (1992) argued that teacher knowledge is gained through experience. It comes in what they called "craft knowledge" (p. 387). They said that craft knowledge was a "form of professional expertise... [representing] the construction of situated, learner-focused, procedural and content-related pedagogical knowledge through deliberate action" (p. 393).

Cochran-Smith and Lytle (1999) extended this work on teacher knowledge by describing three types of knowledge that teachers can bring to bear on their practice, 1) knowledge-for-practice, 2) knowledge-in-practice, and 3) knowledge-of-practice. Knowledge-for-practice was defined as that "given" to teachers by university-based researchers, knowledge-in-practice was embedded in teachers' practice and their reflections on their practice, and knowledge-of-practice was defined as the knowledge gained by teachers when they treat their own classrooms as sites for "intentional investigation" (p. 250).

For this research study, teachers' knowledge-in-practice, or "craft knowledge", and teachers' knowledge-of-practice were emphasized. I observed exemplary teachers' practices of teaching ELLs, and then I interviewed them about why they did what they did. They, and I, used their classrooms for the "intentional investigation" described by Cochran-Smith and Lytle (1999), and we worked together to construct a knowledge base for what it takes to be an exemplary teacher of ELLs.

Understanding Exemplary Teachers

Several researchers have studied the differences between novice and expert teachers (e.g. Berliner, 1994; Porter & Brophy, 1988; Taylor, Pressley & Parson, 2000). Common findings include: 1) experts set priorities, plan, and have goals, 2) experts communicate to their students what is expected of them and why, 3) experts are knowledgeable about their students, 4) experts monitor students' understanding by offering regular appropriate feedback, and 5) experts are reflective and thoughtful about their practice.

While specific practices of exemplary teachers have been the subject of research (e.g. Allington & Johnson, 2000; Block & Mangieri, 2003; McDermott & Rothenberg, 2000), only two studies have focused specifically on effective teachers of ELLs (Garcia, 1991; Tikunoff, 1983). Research on exemplary teachers in general found that these teachers: 1) demonstrated a caring attitude towards and genuine respect for their students, 2) had the traits of perseverance, dedication and enthusiasm, 3) used strategies to actively engage students in the learning process, such as cooperative learning and hands-on activities, 4) reflected often on their practice, 5) demonstrated

culturally relevant teaching, described as knowing the students' lives outside of the boundaries of the classroom walls, and 5) used that knowledge to scaffold students' understanding of curriculum content. Additionally, findings from the Garcia (1991) and Tikunoff (1983) studies on effective teachers of ELLs specifically, showed that these teachers: 1) mediated instruction of ELLs using both the students' native language, Spanish, and English, often alternating between the two for clarification, 2) felt that being bilingual and bicultural would enrich their students' lives, and 3) considered knowing a second language to be an asset.

Effective Practices for ELLs

Researches have identified a common set of effective instructional practices for ELLs, including: 1) providing opportunities for meaningful use of new vocabulary, 2) presenting ideas in both oral and written form, 3) paraphrasing students' remarks and gently encouraging them to expand on their responses, and 4) including questions and activities that require elaborated responses in English so that students can practice expressing their ideas (e.g. Coleman & Goldenberg, 2010; Tellez & Waxman, 2007; Vaughn, Mathes, Linan-Thompson, Cirino, Carlson, Pollard-Durodola, Cardenas-Hagan & Francis, 2006). Echevarria, Short & Vogt (2007) have established a model of instruction aimed specifically at working with ELLs. Their method, the SIOP model (Sheltered Instruction Observation Protocol), provides educators with a comprehensive model for planning and implementing lessons that help ELLs access curriculum content and build language skills.

Methodology

This research set out to examine the backgrounds, beliefs, training and practices of successful teachers of ELLs across contexts – monolingual and bilingual. As such, a multiple, exemplary case study design was appropriate (Yin, 1993). I used constructivist grounded theory (Charmaz, 2000; Glaser & Strauss, 1967) as my method of inquiry. Charmaz (2000) described constructivist grounded theory as "[assuming] the relativism of multiple social realities, [recognizing] the mutual creation of knowledge by the viewer and the viewed, and [aiming] toward interpretive understanding of subjects' meaning" (p. 510). In this study, I and the teachers were mutually constructing a theory of what it takes to be an effective teacher of ELLs; and, how the contexts in which they taught affected their instruction of these students.

Settings

Since one of the goals of the study was to look at the effects of program context on the exemplary teachers and their practices, the study took place at schools with bilingual and monolingual programs, all of which were in the same district. The bilingual program was the only one offered in that particular district. The district was located in an outlying suburb of a major city in the Northeastern part of the United States. Approximately 4,500 students were enrolled in the district. In one of the schools with a monolingual program, meaning all instruction was done in English, the Parker School (all school names are pseudonyms), the percentage of Limited English Proficient (LEP) students was 5.9%. In the second monolingual program, the Washington School, the percentage of LEP students was 12.1%. In the bilingual program, the Lycee School, where instruction was done in both

English and French, the percentage of LEP students was 15%.

Program contexts.

In the bilingual program, all of the students were instructed in both French and English. Students entering the program in Kindergarten receive 90% of their instruction in French and 10% of their instruction in English. Students in Kindergarten are both native-French and native-English speakers. Upon entering first grade, students receive 70% of their instruction in French and 30% of their instruction in English. In second grade, students spend 60% of their time in French, and 40% of their time in English. By third grade, students are instructed 50% of their time in English and 50% of their time in French, and this continues through fifth grade. English-as-a-second language (ESL) instruction was provided for students from Kindergarten through grade five using a "pull out model", meaning ELLs were removed from the classroom approximately 2 to 3 times per week for small group instruction. There were three ESL teachers on staff to serve the ELL students.

In both monolingual programs, all instruction was in English. ESL teachers at all schools used a "pull out" model for working with ELLs, meaning the ELLs were removed from the classroom and instructed in a small-group setting with the ESL teacher. ESL instruction happened approximately 2-3 times per week in each setting for a period of approximately 45 minutes.

Recruitment

The superintendent of an outlying district of a major city was contacted via mail in November of 2005 describing the study and what would be required of the teachers should they agree to participate. In January 2006, principals and administrators in the district were contacted via mail with information about the study. Specifically, the principals and administrators were asked to recommend teachers who were successful at teaching ELLs. The criteria given to the principals and administrators for nomination were: 1) that the teachers had to have been teaching for a minimum of five years, based upon the findings of research on experts and novices (Berliner, 1994) and on exemplary teachers (Allington & Johnston, 2000; Block & Mangieri, 2003), which indicates the important role of experience in developing expertise, and 2) that the nominees' ELLs had to have shown significant progress in their language development in past years on measures used by the school to determine academic progress. These measures included standardized testing scores on state-mandated assessments and holistic writing scores based on the school and district writing prompts given at the beginning and end of the year to all ELL students. However, more criteria were not provided since the major goal of the study was to develop a theory using a constructivist grounded theory design
The principals and administrators who felt they had teachers who could qualify for the study were asked to fill out questionnaires to explain their choices for nominees in detail using a "Qualities of Exemplary Teaching Data Collection Form" (see Appendix A), based on the research done by Block & Mangieri (2003) who collected extensive data on exemplary teachers of literacy. Though the process of nomination can be problematic in selection of participants, research has shown that validity can be maintained by carefully constructing the instruments used to collect the nominations (Allington & Johnson, 2000; Block & Mangieri, 2003; Ladson-Billings, 1994).

Three teachers were nominated from monolingual classrooms: one from Parker School, and two from Washington School. Since the two nominated from Washington School were both first-grade teachers, I decided to choose the teacher whose ELLs had shown greater progress on the school's assessments. Two teachers were nominated by the principal from the bilingual program, and both were chosen in order have two cases who taught in bilingual classrooms and two cases from monolingual classrooms.

Participants

The two teachers from the bilingual program were Rose and Kate (all names are pseudonyms). Rose was a third-grade teacher who had been teaching for over 25 years. She was native-English-speaking, and though not fluent in French, she had a working knowledge of the language through her 25 years of teaching in the bilingual program.

Kate was a fifth-grade teacher who had been teaching for a total of 16 years. It was her third year teaching fifth grade in the bilingual program. She had previously been a middle-school ESL teacher and a high school history teacher. She was fluent in French. During the year of the study, there were four ELL students in Rose's classroom and six ELL students in Kate's classroom. All of the ELL students were French-speakers.

The two teachers from the monolingual programs were Jane and Liz. Jane was a first grade teacher at the Washington School who had been teaching for six years. She had a working knowledge of Spanish, but she said that she did not consider herself fluent in the language. There were six ELL students in her classroom during the year of this study. Four were Spanish-speaking, one was Swedish-speaking and one was Chinse-speaking.

The second teacher, Liz had been teaching for 12 years. She had been both a second- and third-grade teacher, and during the study, she was teaching second grade. She did not consider herself fluent in any language other than English, though she said she had some understanding of Spanish having studied the language in high school. There were five ELL students in her classroom during the year of the study. Three were Spanish-speaking and two were Japanese-speaking.

Data Collection

Data were collected over a period of a full school year (see Table 1). A variety of data were collected including principal questionnaires, interviews with the teachers, observations, stimulated recall sessions, and one end-of-the-year focus group with all four teachers. I did "cluster visits" where I observed three lessons in one week in each teachers' classroom. This way, I was able to see lessons from particular units in various subject areas, allowing for breadth in terms of content areas, and depth of instruction of key ideas and themes across these content areas. Each lesson observed was between 60 and 90 minutes.

Data were collected in seven stages: Stage 1) teacher interviews prior to the beginning of the school year to explore teachers' backgrounds and preparation for teaching, which were tape-recorded and then transcribed by the researcher (Appendix B). The interview questions included ones about the teachers' attitudes towards working with ELLs, what challenges they felt they faced in teaching ELLs, and whether or not they felt it was better for ELLs to be educated in bilingual or monolingual programs and why.

Questions also focused on whether or not they knew how to speak another language, and if they did, how they had learned to do so. In addition, all teachers were asked about where they had grown up and why they had decided to go into teaching.

Stage 2) one cluster visit in each teacher's classroom during their instruction, observing three lessons for each teacher, for a total of twelve lessons. Teachers were audio-recorded using a digital recorder with a wireless microphone the teachers clipped on their shirts while I took field notes to cross-reference these with the audio-recordings. For the field notes, I used a two-column observation protocol (Appendix C). Stage 3) stimulated, audio-recorded recall sessions with each teacher of the observed lessons, during which I took field notes (Appendix D). During the recall sessions, I played the audiotape and stopped it at various intervals to ask questions such as, what were you thinking here when you said that?, or Why did you have your students work in that way? In addition, I would ask the same set of questions each time about their objectives for the lessons, including the planning they did, whether or not they felt the ELLs had understood the objectives of the lesson and why, the assessments they used and why, and any other thoughts they had about how the lesson went.

Stage 4) second teacher interviews to review initial coding and reflect on the initial findings, which were also audio-recorded and during which I took field notes. During this stage, I had created a visual for the participants called "emerging theory of exemplary teaching of ELLs", which included the main categories of "teacher attitude", "background experience" and "classroom practices" with details beneath each category. I asked the teachers to comment on what I had found. My major questions had to do with what they thought of each category and the details supporting that category.
Stage 5) second round of cluster visits, stimulated recall sessions, and discussion of coding and emerging theory, Stage 6) final round of cluster visits, recall sessions and discussion of coding and emerging theory, and Stage 7) focus group to member check emerging theory and reflect upon what it means to be an exemplary teacher of ELLs . During the focus group, the session was audio-recorded and video-recorded, and I asked the teachers four main questions. These were: 1) If you had to prepare a teacher to work with ELLs, what would be the key things you would say you need to know and why?, 2) You teach in different kinds of programs, would you talk about the impact of these programs on your own teaching of ELLs and on the students themselves?, 3) What do you think about teachers' knowledge of students' first language in relation to teaching English to them, do you think it helps or not and why?, and 4) What about the support services in place at your schools for ELLs? Which are helpful for you and for them?

Data Analysis

Data analysis was done in eleven stages (see Table 2). Using constructivist grounded theory, I was analyzing data all the way along the process of collecting data. I had long conversations with the teachers about the data I had collected and my interpretations of the data. I noted what the teachers felt needed to be adjusted and/or expanded upon after each meeting with them, and what they thought was very consistent in terms of how they would interpret the data and who they were as teachers. In this way, as Charmaz (2000) indicated, the teachers and I constructed a knowledge base together of what it takes to

be an exemplary teacher of ELLs and how program context affected their teaching. In addition, I created frequency tables, as indicated in Table 2, for each teacher to show which practices occurred most often among teachers, which occurred more frequently in one context, and which occurred less frequently and then discussed these tables with each teacher during recall sessions and in the final focus group (Appendix E).

Results

The teachers in the study and I identified commonalities across teachers, suggesting a knowledge base of what it takes to be an exemplary teacher of ELLs. These commonalities included four overlapping areas: 1) teachers' knowledge of ELLs, 2) linguistic content knowledge, 3) specific background experiences, and 4) key dispositions.

Teachers' Knowledge of ELL Students

The teachers in this study stressed the importance of "knowing" the ELL students in their class. The areas of knowledge they identified as crucial were: 1) information about their ELLs' previous schooling, 2) what language(s) they spoke at home, 3) the best way to communicate with their students' families, and 4) specific cultural details such as family expectations.

For example, related to the importance of learning about students' prior schooling, Liz realized that that she needed to provide additional information about colonial times for one of her ELL students from Japan. Most of her other students had grown up in the area learning about the American Revolution, whereas this ELL student had no such background knowledge. In understanding this about her student, Liz was able to effectively provide enough background information for the ELL student for him to be able to successfully participate in the unit and lessons.

Jane stated the importance of knowing ELL students' backgrounds in terms of understanding family expectations. She pointed out that she had to learn to communicate solely with a Chinese-speaking ELL student's father when there was any behavioral issue since this was the expectation within that particular student's culture.

Linguistic Content Knowledge

The second area of commonality among the teachers was their linguistic content knowledge (LCK). In all cases, the teachers in this study knew the challenges their ELLs would face in terms of content, and the language within that content. LCK is based on Shulman's (1987) definition of "pedagogical content knowledge" which he defined as "representing the blending of content and pedagogy into an understanding of how particular topics, problems or issues are organized, represented and adapted to the diverse interests and abilities of learners, and presented for instruction" (pp. 15-16). Linguistic content knowledge is distinct from pedagogical content knowledge in that LCK refers to the language needed for ELLs to access content, or "academic language".

The teachers in this study focused specifically on "academic language"; that is, not just on the conversational language that is commonly acquired quickly, but rather, the academic language that is required for students to be successful in school. Cummins (1981) referred to this as the difference between "BICS", basic interpersonal communication skills, or social language, and "CALP", cognitive academic language proficiency", or academic language. Academic language is context-

reduced in nature, and it is a variable that often hinders the academic achievement of ELLs (Fillmore & Snow, 2000; Schleppegrell, 2004). The teachers' LCK was revealed through analyzing data related to their planning and preparation as well as their in-class practices.

Preparation and planning.

The teachers' LCK was demonstrated in their planning and preparation in the following ways: 1) adjusting expectations, 2) expanding ELL students' repertoires for writing, and 3) creating language objectives.

Adjusting expectations.

It is important in defining this category to emphasize that these exemplary teachers did not *water down* curriculum to support their ELLs. Rather, they anticipated what might be particularly difficult for ELLs in terms of language and planned their instruction accordingly. The teachers' adjustments reflected their knowledge of what beginning ELLs would need and what those with more exposure and practice in English would be able to do.

Kate discussed adjusting expectations for her ELLs for a family history project while at the same time not watering down the curriculum. The students in her class all needed to interview a family member. However, she knew from experience that question formation was particularly difficult for her ELL students. Therefore, she scaffolded the assignment for them by providing the correct framing for questions they would need to ask. Her ELL students were doing the same assignment as the other fifth grade students, but with more specific language support planned for and provided by Kate. She said:

Formation of questions is a big issue for ELL students. It's something they always have trouble with, so I provide the frames they need. ELL kids don't necessarily ask that many questions when they're learning language because they tend to be more absorbing of the language and asked to respond to questions rather than formulating them..But with every lesson, the objective for the class as a whole is still the objective for the ELL students...because I'm trying to keep them progressing the way other fifth graders are progressing (second recall session).

Expanding ELL students' repertories for writing.

LCK in the teacher's preparation and planning was also revealed in their focus on expanding their ELLs' repertoires for effective writing. Through careful planning, teachers expanded these repertories in several ways: 1) developing their ELLs' word choice in descriptive writing; 2) helping the ELLs incorporate key content area words into their writing; and 3) aiding ELLs to understand the benefit of planning before writing.

For example, in a writing lesson I observed Jane doing with her first graders, she wanted ELLs to use more sophisticated words to describe objects. She explained in this excerpt how what she called "backwards description" helped build her ELLs' descriptive vocabulary. She explained backwards description as her or a student describing an object and having the other students in the class, especially the ELL students, guess what it was first, then brainstorm additional adjectives to describe the object, and then having the students write the description afterwards. She said:
I have found backwards description to be a way to help ELLs expand upon their

[academic language]. As a newer ELL, you have a limited vocabulary in English, so it can be really difficult to describe something. [Backwards description] is very good for ELLs because they get to hear a lot of different vocabulary before they have to start writing their own descriptions. It gives them more of a foundation to work from (second interview).

Specific language objectives.

The teachers in this study wrote both content and language objectives. These language objectives were more specific language goals for each lesson or unit. To create language objectives for a particular lesson or unit, teachers would make explicit the language demands of the content they needed to teach, and they worked to be sure to focus upon this particular language in their planning so that the ELLs could participate successfully in the lessons.

For example, when Jane was planning a unit on estimation with her first graders, she decided to focus on connecting similar language from other content to aid the ELLs in understanding the word "estimating" in math. To do so, she related the words "prediction" from language arts and "hypothesis" from science to help clarify the idea of estimation in math. Though this benefitted all students, Jane had used these words as specific language objectives for her ELLs. She said,

Relating estimating to what they already knew would not only help all students, but especially the ELLs. I feel like I have to constantly reinforce the language for the ELLs. They can have a hard time with some of these terms, like estimation, so I try to use words they know, like "predicting" to help them understand this new language. (third recall session).

In-class practices.

Teachers' LCK was also revealed through their in-class practices. Two specific practices were: 1) explicitness, and 2) use of ELL students' first language (L1).

***Explicitness*.**

All four teachers were explicit with their ELLs about language errors in their writing and how to correct them. Kate demonstrated explicitness in her teaching. The example below comes from a recall session where Kate talks about the way in which she worked with an ELL student to correct errors in verb tense. Kate emphasized the need to repeat rules in an explicit way in order for the ELLs to eventually be able to correct the errors in their work on their own:

[Helene] made a question with "did" and then used the past tense. That's an extremely common mistake with [ELLs]. I've explained the rule to her at this point multiple times, and she's still doing it, but I just keep correcting her and waiting for it to settle in. And then in her writing journal, if I see the same mistake, I'll circle it and then on the bottom of the page I'll write it out correctly so that when she is writing again, she can look back to remind herself how to fix it (second recall session).

Another example of explicitness comes from Jane. The following excerpt is from my notes during an observation where students were working on their descriptive writing. Here she is working with [Rodrigo], one of her ELLs who is writing about his dinosaur. She works with [Rodrigo] on formulating sentences and correctly using plural forms by using explicit questioning techniques with him:

She goes over to [Rodrigo] to work with him. She helps him formulate his sentences correctly by asking him questions about what he is writing. She helps by writing the sentences directly on the paper. She helps

him with the plural of colors by asking him if there are 1 or 2 colors, and that we put "s" when it has more than one (field notes, second cluster visit).

Use of students' L1.

Using ELL students' L1 demonstrated the teachers' LCK; however, only teachers in the bilingual context used ELLs' L1. Teachers in the bilingual context used the ELLs' native language for support in the following ways: 1) giving translations of words or phrases to clarify concepts, 2) comparing structures in the students' native language to English, 3) providing ELLs with the opportunity to do their writing in their native language.

In an example from Kate's class, she was presenting a social studies lesson in which she was talking with the students about how we learn history. She was doing this as a way to introduce the idea of primary and secondary sources. In this excerpt, Kate is talking to an ELL student, [Helene], who had arrived about a month before this lesson with limited English. This is an example of using ELL students' first language vocabulary to clarify concepts in English, and of allowing ELL students to compose their thoughts and ideas in their native language. Kate says to [Helene]:
What other ways can you learn about history? How do we learn about history? How have you learned history? In your life? (to Helene). Dans ta vie (wait time)….(Helene answers in French)…ok? so write "in class"…and what in class in particular do we learn about history?…where the facts, where the ideas come from…from a book?...Or a teacher? Comment tu apprends l'histoire? D'ou ca vient? (Helene then responds in French)…from your ancestors ? ok, write that down. Do you understand this question? ... Tu comprends?...Then jot down some ideas if you have any ideas…dans cent ans…how could you learn about life in 2007…Comment tu peux apprendre des choses sur maintenant s'il n'est pas des textes d'histoire?...that's sort of the idea, c'est l'idee. Ok?

Specific Background Experiences

Teachers' background experiences that seemed to be influential in the teachers' success with ELLs were: 1) learning a second language, 2) being immersed in a culture other than their own, and 3) years of teaching.

Learning a second language.

All of the teachers in the study expressed that through the experience of learning a second language, whether through school, such as for college requirements (Rose, Jane and Liz), or as part of living abroad later in life (Kate), they were better able to relate to the ELLs' challenges in learning English.

For example, Kate moved to Africa in 1994 with her husband and two small children. They moved to the Cote d'Ivoire, and it was there that she learned to speak French. She did not speak a word of French before she moved there. She said she learned the language by studying it herself and then having to "use it in real life". She emphasized that she had to learn to speak French "not necessarily because she chose to, but because she had to". In speaking with Kate in the two interviews, she was adamant about the fact that her needing to learn a second language was very influential on her attitudes towards working with ELLs and on her ability to teach them effectively. She talked extensively during these interviews about her years in Africa struggling to learn a second language, and how this experience sensitized her to what her ELLs might be

going through when trying to learn English. She talked about "getting upset and exhausted" when trying to communicate in French at the beginning, and that she knew her ELLs felt the same way.

Being immersed in a culture other than their own.

Another background experience that seemed to be influential on these teachers' effectiveness working with ELLs was that of being immersed in a culture other than their own, whether as children or as adults. The teachers articulated that these experiences had positively impacted their work with ELLs in several ways: 1) that they welcomed diversity in their lives and classroom, 2) that they had experienced what it was like to be "the other" and therefore understood what ELLs might be experiencing as "the other" in the classroom, and 3) that they had learned to value differences in people and particularly in their students.

During my first interview with Jane about her background, she described growing up in a diverse community. She felt that this experience not only influenced her ability to teach ELLs, but also her desire to do so. She said:

I lived in a very diverse community growing up, and we had a lot of [ELL children] in our community. So I think I feel like I've always wanted to work with [ELLs] and been used to that in some way. It's what I'm accustomed to. Growing up in my community was really positive for me. I think it made me realize that if [ELLs] are having a difficult time, it's natural. I don't say things like "why can't they learn English?", or "they're not trying to learn", or "why doesn't that family speak English". I like the diversity – it makes things more interesting, so I think I have always felt that way.

In addition, Liz talked about her experience being "the other" when she had traveled abroad to Germany, and she didn't speak the language. She said this had helped her develop empathy for what the ELLs might be going through when they first came to the country not speaking any English. She talked about getting lost on the train in Germany, and how frightening it was for her when she couldn't speak the language and communicate about where she needed to go. She said: "I think if had been thrown into a German school when I was over there, I think I'd be floored. I think it would be so difficult and scary" (second interview).

Jane expressed this sense of "fear" of not knowing the language when she was traveling abroad as well and how this helped her relate to her ELLs. "I think traveling abroad is a really good experience if you are going to be working with [ELLs]. You get a sense of what it is like for [ELLs]. *I was literally afraid* when I didn't understand what people were saying. So I think it kind of gives you that feeling internally of what the trepidation may be like for an [ELL] in your classroom" (second interview).

Years of teaching experience.

All of the teachers in the study had over seven years of teaching experience. These years of teaching experience appeared to have influenced their work with ELLs in the following ways: 1) understanding ELLs' general language development, 2) developing a comfort level in terms of not having a fear or lack of confidence in working with ELLs, 3) developing a "repertoire" of effective strategies for working with ELLs, and 4) understanding that the ELLs in their class were "their

responsibility", not just the ELL teachers' responsibility.

Liz spoke specifically to this notion of responsibility during my second interview with her when we were discussing how her years of teaching ELLs might have impacted her work with these students. She said:
The way I look at it, you're responsible for the development and the growth of the [ELL] child in the classroom, and then I feel like the ELL teachers support what I'm doing in the classroom. It's my responsibility, and I'm being supported by the ELL teacher, and with that I need to advocate for the student. That if they need more support, you get it, and if you think they need less support, and you want more time with the child in the classroom, that I'm advocating that, for what's best for the children.

What is important to stress here, however, was that it was not just the years of teaching that had positively affected these teachers' ability to work with ELLs. It seemed to be their capacity to *reflect* on their working with ELLs over the years that positively affected their ability to teach them. The teachers talked about reflecting on such things as what worked in their lessons and what did not in relation to their ELLs.

Key Dispositions

All four teachers in this study had commonalities in dispositions. Dispositions in this case are defined as tendencies "to exhibit frequently, consciously, and voluntarily a pattern of behavior that is directed to a broad goal" (Katz, 1993, p. 1). The commonalities were: 1) sensitivity, 2) encouragement, 3) positive attitude toward teaching ELLs and, 4) humor about themselves.

Sensitivity.

In all cases, the teachers in this study demonstrated sensitivity towards their ELLs. Sensitivity is defined as the teachers' insight into the psychological and emotional needs of their ELLs. In my two interviews with Rose, she spoke often about these needs. It was clear from our conversations that she thought teachers needed to know how difficult it can be for ELLs when they first arrive. In our first interview, she said:
Sometimes you get [ELL students] who are not happy to be here. Those are the ones who can be really resistant. It can take those kids the whole first year to get comfortable, and it can be miserable for them. They miss everything that is "home", so you can't push it. But you have to support them. Let them know that the next day will be better. You have to consider every day that if they do something in English, it's great; and that gradually, they will succeed (first interview)

Encouragement.

I defined this category as one in which the teachers were able to urge their students on by using words to motivate them, such as "great," "wow," "keep it going," "nice job," among others. In this example from Liz, she is having students read their examples of transitional words and phrases, and she is highlighting on of her ELL's work and encouraging her with words:
[Tomika], you started yours off nicely. Do you want to share it? Great, go ahead. (Tomiko reads the beginning of her story). Good word choice (Tomiko continues reading). Love it!! Love it, and I like [Tomika] did something a little different, she almost told the story from an outside point of view, so very interesting! Great job! (second observation, second cluster visit).

Positive attitude.

The teachers in this study never considered the ELLs in their classrooms to be a burden. The teachers saw having these students as an opportunity to improve their own teaching by trying out different techniques, using flexibility within their lessons, and constantly rethinking units and lessons with ELLs' needs in mind. Kate talked about how ELLs had positively affected her own growth in terms of always wanting to improve her teaching. She said, "Every year I'll think of something that I think I can do a little bit better for [the ELLs], and I'll make a change to make it better" (first interview).

Humor.

The teachers all had senses of humor about themselves which was an endearing quality. I could see that although they were extremely thoughtful and planned when it came to effectively working with ELLs, they also found humor in their failures and successes with these students. They demonstrated the importance of using humor, even at the expense of looking pretty silly, to help their ELLs feel more relaxed in the classroom. I often observed the ELLs, as well as other students in the teachers' classrooms, laughing with the teachers and enjoying themselves. Below, Kate described using humor as a "magnet" for her ELL students

I think [humor] keeps the ELL kids paying attention. I think it attracts them to try to listen to my class more. You know, if the [ELL students] sense that there's something funny going on, they're gonna be more likely to try to want to join into that stream than if I'm just sort of a teacher sitting up there in a very serious dry way. I don't think there's going to be much of a magnet for them to be attracted to trying to follow along (second recall session)

Role of Contexts

The results from this study suggest that teaching in a bilingual context was a more positive experience for the teachers and the ELLs. In terms of the experience for teachers, those in the bilingual context felt it was "easier" for them in terms of instruction. Teachers in both contexts agreed that there were more testing pressures on teachers in the monolingual context; and, that teachers in the bilingual context had an advantage in more quickly determining whether or not ELLs had learning disabilities.

With regard to the experiences of ELLs, teachers in the bilingual context perceived ELLs feeling their first language (L1) was valued, and that ELLs felt less stress overall in the bilingual context.

Ease of instruction.

Kate mentioned several times during the interviews, recall sessions and focus group that she felt it would be harder for her if she needed to teach ELLs in a program where she could not use or did not know ELLs' L1. The best example of this is a quote from the focus group. She said simply:
It makes teaching [ELLs] at our school a lot easier because we have the hook with knowing students' [L1], whereas if you had a multi-language situation [like in the monolingual programs] and you're trying to teach speakers of all types of languages, it would be much harder.

Rose echoed this sentiment during my second interview with her. She said:

It's more difficult for teachers in a monolingual system. I mean, I don't have to

know four languages, I just have to know one. In a monolingual setting you really have to work harder to make sure the ELLs understand. It's never a problem here because you have all the other kids in the class who speak [ELLs' L1]. I think that makes it easier for us.

Testing pressures.

When I asked Liz the question during our first interview about what she thought teachers needed to know to teacher ELLs effectively, she said that they needed to be prepared for the pressures of standardized testing. She talked about how difficult this was for her and for the ELLs in her class since they were required to take the math portion of the state-mandated assessments, even if the ELLs had only just arrived to the school without speaking any English. ELLs in the bilingual program did not have to take these same state-mandated assessments.

Ability to identify learning disabilities.

Jane and Liz expressed in our interviews that not knowing the ELLs' L1 nor having someone readily available to translate put them and the ELL child at a disadvantage. They could not always effectively identify whether an ELL was struggling due to not understanding English or due to a learning disability. In the bilingual program, assessments were available in both French and English, and there was a French learning specialist at the school as well. Therefore, any learning issues could be quickly identified and addressed, providing the ELL with appropriate supports in his or her L1.

Value of L1.

The teachers in the bilingual program felt that having a bilingual program sent a message to the ELLs that their L1 was valued. During our first interview, Kate said, "When [ELLs] first language is used for instruction, it empowers them rather than making them feel like their [L1] is something to be gotten rid of". Rose also felt that the bilingual program showed the value of ELLs' first language and of being bilingual. In our first interview, she said: "I think at our school we recognize that the child's [L1] has value. I think that's very important. If the child knows his [L1] is valued, and important, and looked upon as something positive, then that makes them feel better about themselves, and they perform better. The see that we are making them bilingual, not trying to make them monolingual".

ELLs' stress levels.

There was the perception by all of the teachers that the ELLs in the bilingual program did not feel as much stress as they would if they had to learn English in a monolingual program. When I asked them in the focus group about how they thought the program they taught in affected their teaching and the students, Kate said that because they knew French, it was like "a piece of driftwood to hold onto" for the ELLs. In other words, it wasn't a "sink or swim" situation .

Liz also spoke about how much more quickly she thought her ELL students could adjust if they could use their L1 consistently like in the bilingual program. She said: "I know it's a challenge for [Atsuko] that no one speaks Japanese here. Whereas if everyone were speaking Japanese and could speak Japanese, that it would be easier for him. I think he would feel more comfortable".

Discussion and Implications

In 2005, Bransford, Darling-Hammond and LePage developed a framework to codify a knowledge base for teaching. Their framework included three intersecting areas of knowledge: knowledge of learners and their development within social contexts, knowledge of subject matter and curriculum goals, and knowledge of teaching in light of the content and learners to be taught. The theory developed from the current study for the knowledge needed to become an effective teacher of ELLs (Figure 1) modifies the three areas by Bransford et. al. and adds two additional areas. The modified areas are: 1) knowledge of ELLs (from knowledge of learners and their development) and 2) linguistic content knowledge (from knowledge of subject matter). The two additional areas are: 1) specific background experiences, and 2) key dispositions. In the next section, I will discuss how each modified area and new area relates to current research as well as the implications of these findings.

Knowledge of ELLs

Knowledge of students in effective classrooms for ELLs requires more than just theories of learning and development and the effects of social context on learning (Bransford, et. al., 2005). Knowledge of ELLs speaks specifically to the kinds of information teachers need to gather about their ELLs in order for effective instruction to occur. When teachers do not make attempts to find out about their ELL students' cultures, this can negatively impact teachers' perceptions of their ELLs and potentially negatively impact their ELLs' learning (Echevarria et al., 2007; Huss-Keeler, 1997; Rueda, August, & Goldenberg, 2006).

Teachers can easily be informed about ELLs' background knowledge using guides such as the one provided by Brisk and Harrington (2007) called "Protocol to gather information about learners" (p. 211). Their guide provides teachers with the kinds of questions to ask students' families and to ELLs to get information about their schooling experiences, attitudes about learning English and being in this country, and other personality traits.

Linguistic Content Knowledge

The second category in the theory suggests that effective teachers of ELLs need knowledge of the specific language demands of the content they need to teach. This differs from the Bransford et al. (2005) model in that they emphasized teachers having subject matter knowledge, but they did not discuss the importance of teachers understanding the language demands of the subject matter. Current research suggests that teachers must be able to understand the language-specific demands of content in order to anticipate what language might be problematic for ELLs (Brisk, 2006; Echevarria, Vogt & Short, 2007; Fillmore & Snow, 2000; Schleppegrell, 2004; Tellez & Waxman, 2005; Valdes, Bunch, Snow, Lee, & Matos, 2005). Fillmore and Snow (2000) maintain that with the increasing numbers of ELLs in classrooms, "today's teachers need a thorough understanding of how language figures in education," and that, "too few teachers . . . understand the challenges inherent in learning to speak and read Standard English" (pp. 1-2). In essence, teaching content includes teaching the language required to understand that content.

Specific Background Experiences

The third category suggests that the teachers in the study were effective due to certain background experiences. These experiences were learning a second language and being immersed in cultures different from their own. The role learning a second language in effectively teaching ELLs is consistent with current literature (Baca & Escamilla, 2005; Hyatt & Beigy; Nieto & Rolon, 1997). It seems to "giv[e] [teachers] insight into the language and learning process and the experiences of their students" (Lucas & Grinberg, 2008, p. 611). Youngs & Youngs (2001), in a study of 143 teachers, found that teachers who had completed one or more years of foreign language classes in high school or college were "significantly more positive about teaching [ELL] students than were teachers who had not taken any foreign language classes" (p. 110).

In addition, Youngs & Youngs (2001) found that classroom teachers who had lived or taught outside of the U.S. had significantly more positive attitudes towards working with ELLs than those lacking such experiences. However, of note, they found that simply traveling abroad did not affect attitudes– that it was the importance of an extended period of time that seemed to matter. However, findings from this study suggest that even a shorter period of time did make a positive difference for these teachers.

Key Dispositions

The importance of teachers having certain types of dispositions has been placed at the forefront of teacher education by organizations such as NCATE (National Council for Accreditation of Teacher Education, 2007). The types of dispositions defined as important for teachers to have are in line with the kinds of dispositions found of the effective teachers in this study. For example, NCATE's Standard 1 requires that: Candidates preparing to work in schools as teachers or other school professionals know and demonstrate...professional dispositions necessary to help all students learn. In order to meet the target for "professional dispositions", teacher candidates must "demonstrate classroom behaviors that create caring and supportive learning environments and encourage self-directed learning by all students" (NCATE, 2007).

Part of this notion of a "caring and supportive environment" could be linked to the teachers in this study's use of encouragement. As suggested in research on second language acquisition, motivation can be a critical element to successfully acquiring a second language (Gass & Selinker, 2001; Krashen, 1982; Lightbown & Spada, 1999; Skehan, 1989). In fact, it appears to be the second strongest predictor of success behind aptitude.

In addition, all of the teachers in the study had affirming, positive attitudes towards working with ELLs. This is consistent with research on culturally responsive teaching which emphasizes the critical role teachers' positive attitudes towards their students' and cultures plays in these students' success (deJong & Harper, 2005; Delpit, 2002; Maxwell-Jolly & Gandara, 2002; Villegas & Lucas, 2002). "Teachers who view linguistic diversity and bilingualism as resources rather than deficiencies are also more likely to recognize that limited proficiency in English is not equated with limited ability to learn" (Lucas & Grinberg, 20008, p. 613).

Conclusion and Future Directions

A knowledge base necessary to become an effective teacher of ELLs has been posited

here. This knowledge base includes the four overlapping areas of: 1) teachers' knowledge of ELLs, 2) linguistic content knowledge, 3) specific background experiences, and 4) key dispositions. Though limited in sample size, findings from this study reflect current research suggesting that teacher preparation programs must take these areas into account in order to better prepare teachers to work with ELLs (deJong & Harper, 2005; Lucas & Grinberg, 2008; Tellez & Waxman, 2006). The preparation could take the form of mentors who are particularly effective at working with ELLs, guided experiences interacting with a linguistically-diverse community, explicit instruction in language structures in English, and identifying and building on ELLs' prior knowledge. In addition, findings from this study suggest that teachers in the bilingual program found identifying learning difficulties easier, and perceived that ELLs felt that their language was valued because they were learning through their L1 in addition to English. This suggests that rather than eliminating bilingual programs, consideration should be given to how these programs can help both ELLs and their teachers.

The importance of background experiences and dispositions to effective teaching of ELL students are ones in which more research would be necessary. It would be useful to determine if other effective teachers of ELL students have and the kinds of dispositions of the teachers in this study. In this way, it may be possible to determine if there should be certain requirements of teachers who are going to work with ELL students, such as knowing how to be encouraging and caring towards ELL students. Finally, if ELL students are to succeed in classrooms, regardless of the type of program, further research into the knowledge bases of successful teachers of ELLs is critical.

References

Allington, R. L., & Johnston, P. H. (2000). What do we know about effective fourth grade teachers and their classrooms? Washington, DC: Office of Educational Research and Improvement.

August, D. & Shannahan, T. (2006). *Developing literacy in second-language learners: Report of the National Reading Panel on language-minority children and youth.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baca, L, & Escamilla, K. (2005). Educating teachers about language. In C.T. Adger, C.E. Snow, & D. Christian (eds.), *What teachers need to know about language*, 71-84. Washington, D.C.: Center for Applied Linguistics.

Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161-186). Fort Worth, TX: Harcourt Brace College Publishers.

Block, C. C., & Mangieri, J. (2003). *Exemplary literacy teachers: Promoting success for all children in grades K-5*. New York: The Guilford Press.

Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling- Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 1-39). San Francisco: Jossey Bass.

Brisk, M. & Harrington, M. M. (2007). *Literacy and bilingualism: A handbook for ALL teachers (2nd ed).* Mahwah, NJ: Lawrence Erlbaum

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). Thousand Oaks, CA: Sage Publications, Inc.

Cochran-Smith, M. & Lytle, S.L. (1999). Relationships of knowledge and practice: Teacher learning in communities. In A. Iran-Nejad & C.D. Pearson (Eds.). *Review of research in education* (pp. 249-305).

Coleman, R. & Goldenberg, C. (2010). What does the research say about effective practices for English Learners? Part III: Promoting literacy development. *Kappa Delta Pi Record, 46*(3), 106-11.

Coyne, M., Kame'enui, E. & Carnine, D. (2011). Effective teaching strategies that accommodate diverse learners (4th ed.). Upper Saddle River, NJ.

deJong, E. & Harper, C. (2005). Preparing mainstream teachers for English-Language learners: Is being a good teacher good enough? *Teacher Education Quarterly*, Spring.

Echevarria, J., Vogt, M., & Short, D. J. (2007). *Making content comprehensible for English language learners: The SIOP model*. Boston: Allyn & Bacon.

Facella, M. A., Rampino, K. M., & Shea, E. K. (2005). Effective teaching strategies for English language learners. *Bilingual Research Journal, 29*(1), 209-221.

Fillmore, L. W., & Snow, C. (2000). *What teachers need to know about language* (No. ED-99-CO-0008). Washington, DC: Center for Applied Linguistics.

Garcia, E. (1991). Effective instruction for language minority students: The teacher. *Journal of Education, 173*(2), 130-141.
Garcia, E., & Curry-Rodriguez, J. (2000). The education of limited English proficient students in California schools: An assessment of the influence of Proposition 227 on selected districts and schools. *Bilingual Research Journal, 24*(1 & 2), 15-35.

Graves, A. W., Gersten, R., & Haager, D. (2004). Literacy instruction in multiple-language first-grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research & Practice, 19*(4), 262- 272

Grimmett, P. & MacKinnon, A. (1992). Craft knowledge and the education of teachers. In *Review of research in education* (vol. 18, pp. 385-456). Washington, D.C.: American Educational Research Association.

Hyatt, D.E. & Beigy, A. (1999). Making the most of unknown language experience: Pathways for reflective teacher development. *Journal of Education for Teaching, 25*(1), 31-40.

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children.* San Francisco: Josey-Bass.

Lucas, T. (2005). Fostering a commitment to social justice through service learning: Hopes, plans, and realities in a teacher education course. In N.M. Michelli & D.L. Keiser (eds.), *Education for democracy and social justice* (pp. 167-188). New York: Routledge.

Lucas, T. & Grinberg, J. (2008). Responding to the linguistic reality of mainstream classrooms: Preparing all teachers to work with English language learners. In Cochran-Smith, M., Feiman-Nemser, S., McIntyre, D.J. & Demers, K.E. (Eds.). *Handbook of Research on Teacher Education, 3rd Edition*. (pp. 606-636). New York, NY: Routledge

Mathes, P.G., Pollard-Durodola, S. D., Cardenas-Hagan, E., Linan-Thompson, S., & Vaughn, S. (2007). Teaching struggling readers who are native Spanish speakers: What do we know? *Language, Speech and Hearing Services in Schools, 38,* 260-271.

McDermott, P.C. & Rothenberg, J.J. (2000). *The characteristics of effective teachers in high poverty schools – Triangulating our data.* Paper presented at the American Educational Research Association, New Orleans, LA.

Miles, M. B., & A. M. Huberman (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA, Sage.

Moll, L., Amanti, C., Neff, D. & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice, 31*(2), 132-141.

Palmer, D.K. & Garcia, E. (2000). Voices from the field: Bilingual educators speak candidly about Proposition 227. *Bilingual Research Journal, 24*(1&2), 169-178.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. *Educational Leadership, 45*(8), 74-85.

Reyes, L. L. (1992). Challenging venerable assumptions: Literacy instruction for linguistically different students. *Harvard Educational Review, 62*(4), 427-446.

Ruddell, R. B. (1997). Researching the influential literacy teacher: Characteristics, beliefs, strategies, and new research directions. In C. K. Kinzer, K. A. Hinchman, & D. J. Lue (Eds.), *Inquiries in literacy theory and practice: 46th yearbook of the National Reading Conference* (pp. 37-53). Chicago: National Reading Conference.

Rueda, R. S., August, D,. & Goldenberg, C. (2006). The sociocultural context in which children acquire literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second language learners: Report of the national literacy panel on language-minority children.* Erlbaum Associates.

Schleppegrell, M. J. (2004). *The language of schooling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shanahan, T., & Beck, I. (2006). Effective literacy teaching for English Language Learners. In D. August & T. Shanahan (Eds.), *Developing literacy in second language learners: Report of the national literacy panel on language-minority children and youth* (pp. 415-488). Mahwah, NJ: Lawrence Erlbaum Associates.

Stritikus, T. & Garcia, E. (2000). Education of limited English proficient students in California schools: An assessment of the influence of Proposition 227 on selected teachers and classrooms. *Bilingual Research Journal, 24*(1&2), 75-85.

Taylor, B. M., Pressley, M., & Pearson, D. (2000*). Effective teachers and schools: Trends across recent studies.* Washington, DC: Office of Educational Research and Improvement.

Tellez, K., & Waxman, H. C. (2005). *Effective professional development programs for teachers of English Language Learners.* Retrieved July 19, 2008, from www.temple.edu/lss

Tellez, K. & Waxman, H.C. (Eds). (2006). *Preparing quality educators for English language learners.* Mahwah, NJ: Lawrence Erlbaum

Tikunoff, W. J. (1983). *Compatibility of the SBIF features with other research on instruction for LEP students* (Research/Technical No. SBIF-83-R.9/10). San Francisco: Far West Lab for Educational Research and Development.

U.S. Department of Education, National Center for Education Statistics (2011). *The Condition of Education 2011* (NCES 2011-033).

Valdes, G., Bunch, G., Snow, C., Lee, C., & Matos, L. (2005). Enhancing the development of students' language(s). In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 126-168). San Francisco: Jossey-Bass.

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola,S., Cardenas-Hagan, E. & Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*(2), 153-180.

Villegas, A.M. & Lucas, T. (2002). *Educating culturally responsive teachers: A coherent approach*. Albany, NY: SUNY Press.

Wiggins, R. A., & Follo, F. (1999) Development of knowledge, attitudes, and commitment to teach diverse student populations. *Journal of Teacher Education* v. 50 no. 2 (March/April) p. 94-105.

Wong Fillmore, L. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly, 6,* 323-346.

Youngs, C. S. & Youngs, G.A. (2001). Predictors of mainstream teachers' attitudes toward ESL students. *TESOL Quarterly, 35*(1), 97-120.

Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms, grades 5-12.* San Francisco: Jossey -Bass.

TABLE 1
Data Collection Timeline

Table1
Data Collection Timeline

Types of Data	Aug06	Sept06	Oct06	Nov06	Dec06	Jan07	Feb07	Mar07	Apr07	May07	June07
Interviews	x						x				
Observations		x	x		x	x		x	x		
Recall Sessions				x			x			x	
Field Notes/ Memos	x	x	x	x	x	x	x	x	x	x	x
Focus Group											x

Table 2
Data Analysis Stages and Descriptions

Stage 1: Memoing	Memoing was done while transcribing all observations, interviews and recall sessions. During this process, initial themes and patterns among the data were noted. Memos began with simple words and phrases, such as "compassion", "laughter", "questioning ELLs", "slow and deliberate speed when talking", etc. A time stamp was placed next to each phrase or word.
Stage 2: Open Coding	Initial coding, or open coding (Strauss & Corbin), was done while memoing. The open codes were not placed into more specific categories until second round of observations and recall sessions were completed.
Stage 3: Data Display	After transcribing the second round of observations and recall sessions, data were categorized into larger themes and ideas and put into a data display (Miles and Huberman, 1994).
Stage 4: Initial Frequency Tables	Under each category in the initial data display, instances from the time-stamped data were plugged in to note the frequency of occurrences.
Stage 5: Initial Diagrams	Visuals were created called "emerging theory of exemplary teaching of ELLs" and "contextual influences on teaching ELLs" and presented to the teachers for input/comment.
Stage 6: Refining Theory	New data from recall sessions, interviews with teachers, and observations were used to refine theory. Codes were refined for teacher practices to include "preparation and planning", "in-class practices" and "observable attitudes towards ELLs".
Stage 7: Focus Group	All teachers met together at end of the school year to member-check the refined theory of what it takes to be an exemplary teacher of ELLs.
Stage 8: Line by Line Axial Coding	Line by line analysis of each teacher's interviews and field notes during the interviews looking for any additional details about their background experiences that may have influenced their successful teaching of ELLs.
Stage 9: Nvivo Analysis	To further examine practices the teachers used, all data was entered into Nvivo and tree codes were developed for the main categories of "preparation and planning", "in-class practices" and "observable attitudes". Sub-codes were added under each main code such as "use of visuals" under in-class practices". Coding reports were run for each code indicating the sentences, phrases or words that fell under each sub-code for each teacher.
Stage 10: Final Frequency Tables	After analyzing the coding reports, frequency tables were made (Appendix E) using main categories and sub-codes for each teacher to show which practices occurred most often among teachers, which occurred more frequently in one context, and which occurred less frequently.

Appendix A

Qualities of Exemplary Teaching, Data Collection Form

Directions: The purpose of this study is to identify the qualities possessed and regularly exhibited by exemplary teachers of English language learners. You have been selected because of your expertise as a supervisor of these teachers. The information of the first page will remain confidential. The answers that you provide will be analyzed anonymously by the researcher. When all data have been tallied, the researcher will ask you to confirm the accuracy of the data and to change findings that were interpreted inaccurately.

Part 1

Name _____

Title _____

School District _____

Address _____

City, State, Zip Code _____

Phone number _____

Email _____

Fax _____

Number of years of supervisory experience _____

Number of years at current position _____

Grade levels that you supervise _____

Name and grade level of the teacher you are nominating _____

Part II

Reflect on whom you have identified as an exemplary teacher of English language learners. Select the two most important behaviors that distinguish this exemplary teacher in his or her abilities to teach English language learners. State two characteristics that make this teacher effective at teaching English language learners at his/her particular grade level. These characteristics are the one that you most credit this teacher's success in teaching English language learners. Your descriptions can be written in a global, all-inclusive format or can be written in a more specific manner. Please write your responses below:

Appendix B
First Interview Protocol

Question	Designed to Find Out….
Where are you from?	Teacher background; put teacher at ease
How long have you been teaching?	Teacher background; specific information
What was your first teaching job?	Teacher background; reflect on past experiences
Why did you go into teaching?	Teacher background; motivation
Do you know how to speak another language? If so, how did you learn it?	Teacher background
What has been your training in working with English language learners?	Teacher background; specific information on studied population
How many ELLs have you had in your classes over the years?	Teacher background; specific information on studied population
What kinds of strategies do you use with the ELLs in your classroom? Are they different from what you use with your regular mainstream students?	Teaching methods; specific information on studied population
Can you give me some examples of what strategies you use to teach language specifically?	Teaching methods; specific information on language teaching
Have you been trained to teach language specifically?	Teaching methods; specific information on language teaching
How do you think ELL students acquire English? Do you have a theory of second language acquisition?	Teacher knowledge; reflection on language theory
How do you feel about have ELLs in your classroom?	Teacher understanding; reflection on specific population
Have your feelings changed about these students over the years? If so, how have your feelings changed?	Teacher understanding; reflection/inquiry on specific population
What are some of the biggest challenges you have faced in teaching ELLs? Have these challenges changed over the years?	Challenges teachers face in each setting
Have you received additional support in teaching ELLs? If so, what kind of support have you been given?	Challenges teachers face; kind of support that is in place in each setting; teacher training
Are there any kinds of services you feel the ELL students in your classroom should be getting that they are not getting?	Teacher support/challenges; information on studied population
What are some of the biggest successes you have had in teaching ELLs?	Teacher background; successes teachers have had with studied population
What would you say teachers need to know in order to teach ELLs successfully?	Specific information on teaching studied population

How would you say your setting of a bilingual/monolingual English immersion environment influences how you teach ELLs?	Contextual influences on teaching ELLs
How do you feel about bilingual education?	Teacher attitudes
Do you feel it is better for ELL students to be educated in English only immersion or have a bilingual education? Why?	Teacher attitudes
How do you feel it might be for the ELL students in your class to be learning English?	Teacher attitudes
What do you think makes you so effective in working with ELLs?	Teacher beliefs; teacher inquiry
Are there other teachers in the school who you think are effective at working with ELLs? Why do you think they are effective?	Teacher beliefs; teacher inquiry
What is your relationship with the parents of the ELL students in your classroom? Do you communicate with them? If so, how?	Teacher attitudes; information on studied population
Do you feel that the ELL parents are involved in their students' education? How or how are they not?	Teacher attitudes; information on studied population
What are some of the biggest challenges the ELL students in your classrooms face?	Challenges in each setting
How do you help them with these challenges?	Challenges in each setting; teacher training
What else do you feel is important to know about your working with ELLs?	Additional information; teacher reflection
Are there other things you would like to talk about with regard to your teaching ELLs that we haven't addressed yet?	Additional information; teacher reflection

Appendix C
Simulated Recall Session Protocol

Teacher Name:
School:
Lesson Recalled:
Date of Lesson:
Today's Date:
Time:

*Notes were typed in as the teacher recalled the lesson. Recall session followed this general outline.

What was your objective for the lesson?	
Why did you use this particular strategy?	
What aspect of language were you hoping to teach in this lesson?	
What was the target for the ELL students during this lesson?	
Can you tell me what you were thinking when you began the lesson?	
What was the procedure that you were going through in your mind?	
Why did you have your students work in this way? (group or individual work)	
Did you feel that the students understood the lesson?	
What tools did you use for assessment?	
What challenges do you feel the ELL students had during this lesson?	
Did you feel that the lesson went well? Why or why not?	

Appendix D
Observation Protocol

Teacher Name:
School:
Date:
Lesson Observed:
Location:
Time:

Action Observed	Interpretation

Appendix E

Number of Occurrences: Preparation and Planning

Categories	Kate bilingual context	Rose bilingual context	Liz monolingual context	Jane monolingual context
A. Preparation/Planning				
1. Use of themes/units	42	41	27	27
2. Making connections among concepts	53	44	33	51
3. Knowledge of students	93	96	106	101
4. Adjusting expectations	72	85	68	71
5. Grouping/pairing (in planning phase of lesson)	28	27	22	28
6. Comfortable environment	38	48	37	25
7. Focus on academic language[b]	63	50	121	99
8. Language objectives[b]	66	66	119	98
9. Word rich environment[b]	45	44	79	90
10. Organization[a]	64	57	123	89
11. Exemplars/models[a]	44	39	76	40

[a]Occurrences that were higher for Liz

[b]Occurrences that were much higher in certain contexts

Number of Occurernces: In-class Practices

Categories	Kate bilingual context	Rose bilingual context	Liz monolingual context	Jane monolingual context
B. In-class Practices (during lessons)				
1. Use of visuals[a]	62	71	60	69
2. Explicitness[a]	66	73	88	77
3. Frequent check-ins with ELL students[a]	94	74	105	81
4. Use of first language[b]	49	32	6	8
5. Repetition of key vocabulary and phrases[b]	64	52	119	86
6. Routines[b]	18	15	59	55
7. Prompting/coaching[c]	76	83	162	90
8. Use of exemplars/models[c]	63	56	93	42
9. Wait time	35	47	43	37
10. Gestures	11	10	0	4
11. Rephrasing	31	32	22	30
12. Making connections	37	42	50	50
13. Slow speech	21	29	35	19
14. Humor	26	27	15	13
15. Grouping	27	22	32	22
16. Selective about corrections in writing	16	19	16	22

[a]Occurrences that were the highest for all teachers
[b]Occurrences that were higher in certain contexts
[c]Occurrences that were higher for Liz

How Teachers' Perceive Principal Supervision and Evaluation in Eight Elementary Schools

Author Info:

Bret G. Range is an Assistant Professor at the University of Wyoming. Jason Anderson is Director of Elementary Operations at Springfield Public Schools. David Hvidston is an Assistant Professor at the University of Wyoming. Ian Mette is an Assistant Professor at the University of Wyoming.

Correspondence regarding this manuscript should be addressed to Bret G Range, University of Wyoming, Education Room 310, 1000 E. University Ave, Laramie, WY 82071. Ph. (307) 766-6825, Email: brange@uwyo.edu.

Abstract

The purpose of this study was to understand teachers' perceptions regarding principals' supervision and evaluation in eight high-performing elementary schools. An online survey was sent to teachers in select elementary schools and findings suggest principals engaged in all seven supervision and evaluation constructs measured, with teachers rating the concept of differentiated supervision and evaluation the lowest. Additionally, there were notable differences in how novice teachers viewed the seven supervision and evaluation constructs when compared with more experienced teachers. Analysis of open-ended items suggested teachers had high levels of trust in their principals based on positive feedback and routine observations, which increased feelings of shared instructional leadership within the schools.

Keywords: formative supervision, summative evaluation, principal supervision, differentiated supervision, and evaluation

Introduction

The role of effective school principals increasing teachers' instructional capacity by routinely visiting classrooms is well documented (Ikemoto, Taliaferro, & Adams, 2012; Ing, 2009; Marzano, 2012; Range, Young, & Hvidston, 2013). The process of visiting teachers' classrooms, providing feedback, and assessing teachers' performance based on multiple measures is collectively referred to as teacher supervision and evaluation (Zepeda, 2013). However, the endgames of supervision and evaluation have conflicting outcomes causing confusion between educators. Supervision is used to collect multiple data points concerning teachers' performance with the goal of improving instructional abilities (Hinchey, 2010). Evaluation is used to assign ratings to teachers' overall performance and is used to determine if teachers' have met minimum benchmarks (Glickman, Gordan, & Ross-Gordan, 2005). Despite these differences, school districts typically treat both processes as the same, resulting in supervision and evaluation practices which are arduous and lack differentiation based on teachers' various needs (Jacob, Vidyarthi, & Carroll, 2012). With little or no differentiation, all teachers are rated satisfactory, excellent teachers are unrecognized, and early career teachers are not provided special supports (Weisberg, Sexton, Mulhern, & Keeling, 2009).

One way to understand how effective principals connect supervision and evaluation, and differentiate the process based on the needs of teachers is to highlight principals' behaviors in high-performing schools. For example, Ikemoto et al. (2012) analysis of high-performing principals metaphorically categorized them as "playmakers" as they supported and held teams of teachers accountable. Adopting a similar view, whereby principals act as catalysts for instructional excellence and accountability, the purpose of this study was to understand teachers' perceptions in eight high-performing elementary schools concerning their principals' supervisory and evaluative behaviors, and highlight how teachers' perception of principals could strengthen such behaviors.

Supervision and Evaluation

Supervision and evaluation procedures in most school districts position principals to undertake "a conflicting role of supporter and summative judge" (Peterson, 1995, p. 214). These roles have the potential to cause fear and mistrust amongst teachers, especially when the outcomes of both processes are not explicitly explained to teachers (Range, Scherz, Holt, & Young, 2011). Effective principals understand this dilemma and intertwine formative supervision and summative evaluation into a seamless process, continually focusing on teacher growth. For the purpose of the study, principals' supervisory and evaluative behaviors include establishing high standards for teachers' performance, conducting routine classroom observations, providing quality feedback to teachers, using data collected during the supervisory process to inform teacher evaluation, and differentiating supervision and evaluation for teachers based on developmental levels (Derrington, 2011; Donaldson & Donaldson, 2012; Ikemoto et al. 2012; Range et al., 2011; Zepeda, 2013).

Effective principals match their supervision and evaluation practices to the ability levels of teachers, which includes teachers' strengths, weaknesses, and professional development needs, termed differentiated supervision and evaluation (Gupton, 2010; Zepeda, 2013). As a result, the onus is on

principals to diagnose teachers' needs and apply appropriate supervisory and evaluative responses based on those needs. Specifically, principals should be cognizant of three teacher differentiation considerations as they apply supervision and evaluation, including teachers' tenure status, teachers' years of teaching experience, and teachers' job assignment.

First, teachers' tenure status and years of teaching experience are two variables that are closely related. Non-tenured teachers, deemed novice teachers, present a unique challenge for principals as they apply supervision and evaluation (Robertson, 2006). Because non-tenured teachers have limited teaching experience, usually less than three years, novice teachers require considerable guidance in how to overcome low level teaching issues, such as the physical layout of classrooms, materials distribution to students, and classroom management (Cuddapah & Burtin, 2012). Zepeda (2013) has written extensively concerning supervising non-tenured and early career teachers, highlighting their unique needs through the theoretical framework of adult learning. She reports non tenured, early career teachers are in survival mode, in which they seek affirmation from supervisors and a majority of their time is spent on managing unfamiliar situations.

Conversely, effective principals also understand the unique needs of experienced teachers, those with typically more than three years of experience (Jacob et al., 2012). Such teachers are concerned less with low-level teaching competencies and focus more on meeting the needs of students, as well as their own professional growth (Zepeda, 2013). However, adult theorists postulate that many teachers regress during later stages of their teaching careers and exhibit complacency. A variable in this career regression might be teacher supervision and evaluation procedures that do not differentiate to meet experienced teachers unique needs (Weisberg et al., 2009). Experienced teachers require less compliance-driven supervision and desire supervision that builds their capacity in areas of personally identified professional development (Zepeda, 2013).

Finally, a differentiated need not typically highlighted in the literature is principals' supervision and evaluation of content area teachers, usually music, art, physical education, counseling, and special education at the elementary level. The pedagogical skills of these teachers can still be assessed as many of their teaching aptitudes are similar to the proficiencies necessary in regular classrooms. However, as Nolan and Hoover (2008) posited, principals must also possess "content expertise in order to make a fair, informed, comprehensive assessment of teacher performance and competence" (p. 13). As a result, if principals do not have a content specific background or take the time to understand course content within a support area, supervision and evaluation of support teachers is challenging and unfair when assessing good teaching performance.

Context of the Study

The study was conducted in a large, urban Midwest school district which served approximately 24,000 students. In this state, teachers were considered non-tenured until they taught successfully for five consecutive years in the same school district. School districts within the state were given autonomy to adopt teacher supervision and evaluation procedures that met their site-specific needs. The school district selected in this study created its teacher supervision and evaluation procedures by including both

principals' and teachers' feedback on the formative and summative evaluation tools. Non-tenured teachers received one formal and one informal observation by principals each year until they reached their tenure year. Tenured teachers received one informal observation each year and one formal observation by principals every five years. Formal observations consisted of a pre-observation conference, an extended observation, and a post-observation conference. Principals had little autonomy to differentiate supervision for teachers as the same forms (pre-observation conference, post observation conference, and professional development) were used for both non-tenured and tenured teachers. Using student performance data over a three-year period, eight elementary sites were purposively sampled because they were considered high performing by the school district. These elementary schools were considered high performing because their third- and fourth-grade communication arts and math student assessment scores on both the state assessment and district assessment placed them within the top 10% of all elementary schools (*N*=39) in the district. Table 1 displays general demographic information for the eight elementary schools including total school enrollment, percent of students on free or reduced lunch, total number of certified teachers, and teaching backgrounds of principals. Numbers have been assigned to each school to protect their identities and the identity of the school district.

Table 1

Demographic Information for Elementary Schools

School	Enrollment	Free/Reduced Lunch %	Number of Teachers	Background of Principal
1	534	32.2	30	PE teacher
2	518	17.2	30	Classroom teacher
3	346	27.7	21	Classroom teacher
4	467	23.6	29	Music teacher
5	243	48.6	15	Classroom teacher
6	152	54.5	8	Classroom teacher
7	355	24.8	20	Classroom teacher
8	399	26.8	26	Classroom Teacher

Method

The study was descriptive and inferential and used an online instrument to ascertain the attitudes of teachers about their principals' supervisory and evaluative behaviors. As a result, two research questions guided the study: (1) What are elementary teachers' perceptions about principals' supervisory and evaluative

behaviors?, and (2) What are elementary teachers' views about principals' supervisory and evaluative strengths, and how do teachers suggest principals might improve their supervisory and evaluative practices? The online instrument was e-mailed to all 179 certified teachers in the eight elementary schools in April 2013 and resent to non respondents an additional two times. As a result, 74 teachers responded to the survey prior to the survey being closed in May 2013, a response rate of 41%.

Instrument

The instrument used in data collection was a survey used in a previous principal supervision and evaluation inquiry (Clark, 1998), and contained three primary sections. The first section included 22 Likert scaled items (1 = *strongly disagree* to 4 = *strongly agree*) designed to measure teachers' perceptions about seven constructs regarding their principals' supervisory and evaluative behaviors which included (a) establishes standards for teachers' performance (two items), (b) conducts routine observations (two items), (c) conducts adequate pre-observation conferences (four items), (d) conducts adequate post-observation conferences (six items), (e) assesses total performance evaluations (three items), (f) differentiates supervision and evaluation (two items), and (g) has the ability to supervise and evaluate (three items).

To establish internal reliability on the instrument, Chronbach's alpha coefficient was calculated on all 22 scaled items and found to be 0.98. Additionally, Chronbach's alpha coefficients were calculated on each of the seven constructs which included establishes standards for teachers' performance expectations (0.79), conducts routine classroom observations (0.86), conducts adequate pre-observation conferences (0.96), conducts adequate post-observation conferences (0.96), assesses total performance evaluations (0.91), differentiates supervision and evaluation (0.76), and has the ability to supervise and evaluate (0.88). The second section of the instrument included two open-ended items which asked teachers to describe their principals' greatest strengths in supervising and evaluating their teaching performance, as well as ways in which principals might improve these practices. The final section of the instrument collected demographic information on the sample and included gender of the teacher, job assignment, total years of teaching experience, and tenure status.

Data Analysis

Quantitative data were analyzed descriptively and inferentially. Descriptive means were calculated for the entire sample and then means were broken down by tenure status (two groups), years of teaching experience (three groups), and job assignment (two groups). Inferential statistics included independent samples *t* tests for the variables of tenure status and job assignment, and analysis of variance (ANOVA) for years of teaching experience. For the two-open ended items, respondents' answers were open-coded to yield patterns and themes (Jones, Torres, & Arminio, 2006), and coding was characterized as flexible and expandable as themes emerged. After initial coding of the two opened-ended questions by one researcher, the other researchers coded answers to ensure reliability, and researchers checked codes until agreement was reached.

Results

All 74 respondents were female and had an average of 12.67 years of total teaching

experience. More specifically, when looking at teachers' years of teaching experience, 12 (16.2%) had one to three years experience, 17 (23%) had five to 10 years teaching experience, and 38 (51.4%) had 11 or more years of teaching experience. Regarding job assignment, 50 (67.6%) respondents identified themselves as regular classroom teachers (K-5) while 17 (23%) were classified as support teachers (art teachers, music teachers, physical education teachers, special education teachers, or counselors). Finally, 50 (67.6%) respondents were tenured while 18 (24.3%) respondents were not tenured.

The first research question, "What are elementary teachers' perceptions about principals' supervisory and evaluative behaviors?" yielded descriptive findings that help analyze common traits found among the eight principals studied. First, overall means were created for each of the seven supervision and evaluation constructs. Table 2 displays the means for the seven supervision and evaluation constructs based on responses from the overall sample, for tenure status, for teaching experience, and for job assignment. A significance level of 0.05 was used for independent samples *t* tests and the ANOVA.

Table 2

Teachers' Perceptions about Principals' Supervisory and Evaluative Behaviors based on Tenure Status, Years of Teaching Experience, and Job Assignment

	Overall Mean	Tenure Status Mean		Years of Teaching Mean			Job Assignment Mean	
		Tenured (*n*=50)	Non (n=18)	1 to 3 (*n*=12)	5 to 10 (*n*=17)	11 or more (*n*=38)	Classroom (*n*=50)	Support (*n*=17)
Principal has ability to supervise and evaluate	3.40	3.37	3.52	3.42	3.39	3.45	3.37	3.47
Principal conducts routine observations	3.39	3.35	3.64	3.63	3.44	3.39	3.40	3.47
Principal establishes standards for teachers' performance	3.39	3.31	3.61	3.50	3.24	3.46	3.31	3.59
Principal conducts adequate pre-observation conferences	3.37	3.31	3.60	3.58	3.35	3.38	3.33	3.51
Principal conducts adequate post observation conferences	3.37	3.33	3.53	3.56	3.30	3.43	3.33	3.52
Principal assesses total performance evaluations	3.29	3.21	3.56	3.53	3.27	3.28	3.26	3.39
Principal differentiates supervision and evaluation*	3.00	2.95	3.24	3.29	2.59	3.17	3.05	2.94

Note: Scale ranges from 1=*strongly disagree* to 4=*strongly agree*; Support includes related arts teachers, special education teachers, and counselors; * indicates a significant difference at the 0.009 level between teachers with 1 to 3 years of experience and teachers with 5 to 10 years of experience

Overall, teachers agreed with all constructs as all had means higher than 2.50. Teachers strongly agreed most that *principals had adequate abilities to effectively supervise and evaluate* teachers (*M*=3.40), however they strongly agreed least that *principals differentiated supervision and evaluation* for teachers based on need (*M*=3.00). When looking at data through the lens of tenure status, tenured teachers strongly agreed most that *principals had adequate abilities to supervise and evaluate* teachers (*M*=3.37), while non-tenured teachers strongly agreed most that *principals conducted routine observations* (*M*=3.64). Both tenured teachers (*M*=2.95) and non-tenured teachers (*M*=3.24) strongly agreed least that *principals differentiated supervision and evaluation* for teachers based on need. Although there were no significant differences between the attitudes of tenured teachers and non-tenured teachers on any of the constructs, non-tenured teachers agreed more than tenured teachers on all seven constructs concerning their principals' supervision and evaluation skills.

When data were broken down by years of teaching experience, teachers with one to three years of teaching experience (*M*=3.63) and teachers with five to 10 years of teaching experience (*M*=3.44) strongly agreed most that *principals conducted routine observations*. Teachers with 11 or more years of experience strongly agreed most that *principals established standards for teachers' performance* (*M*=3.46). All three groups of teachers (one to three, *M*=3.29; five to 10, *M*=2.59; and 11 or more, *M*=3.17) strongly agreed least that *principals differentiated supervision and evaluation*. Finally, an ANOVA revealed a significant difference between how teachers with one to three years of teaching experience viewed *principals' differentiated supervision and evaluation* when compared to teachers with five to 10 years of experience, $F(2, 63) = 5.11$, $p = 0.009$. Specifically, the less experienced teachers strongly agreed more that *principals differentiated supervision and evaluation*.

Finally, looking at the data through the lens of teachers' job assignment, regular classroom teachers (K-5) strongly agreed most that *principals conducted routine observations* (*M*=3.40). Support teachers strongly agreed most that *principals established standards for teachers' performance* (*M*=3.59). Both regular classroom teachers (*M*=3.05) and support teachers (*M*=2.94) strongly agreed least that *principals differentiated supervision and evaluation*. Finally, support teachers agreed with six of the seven supervision and evaluation constructs more than regular classroom teachers. The only construct in which this was not the case was *principals differentiated supervision and evaluation*, which regular classroom teachers agreed more than support teachers. However, there were no significant differences between regular classroom teachers and support teachers views on any of the supervision and evaluation constructs.

To answer the question regarding elementary teachers' views about principals' supervisory and evaluative strengths, and how teachers suggest principals might improve their supervisory and evaluative practices, responses to the two open-ended questions were open coded by the researchers. Regarding principals' strengths concerning supervisory and evaluative behaviors, 52 teachers responded to this question. Initial coding resulted in nine broad themes and these were condensed into three specific themes, namely *relationships/trust building*, *positive feedback*, and *routine observations*. First, teachers indicated these eight principals

spent considerable time building a relationship with them which created a high level of trust between principals and teachers. This included allowing teachers input into decisions, empathetic listening, and treating teachers as part of a team. Reflecting on building trusting relationships, one respondent stated, "She listens and is very quick to solve problems, she involves teachers in problem solving which helps build trust between teachers and administrators. She is very supportive of how I teach." Another respondent commented, "I feel the greatest strengths are that my principal treats teachers as part of a team. Each teacher is treated as a professional who is contributing to the education of students."

Regarding *positive feedback*, teachers indicated that when principals followed up their classroom visits with feedback, the feedback was framed first with positives before suggestions or reflection questions were offered to teachers. For example, one teacher believed her principal took additional time to provide positive feedback after all classroom walkthroughs. Another respondent remarked, "I know exactly where I stand with my principal on teaching. Ideas are always suggested and my strengths are commended."

The final theme that surfaced was *routine observations*. Teachers indicated these eight principals were in classrooms daily which provided them a clear understanding of teaching and learning issues in their schools. One respondent observed, "With weekly walkthroughs, my principal gets a chance to see what is going on consistently in my room." Another teacher believed, "My principal is in my classroom multiple times a week, so I feel my principal gets an accurate overall picture of my classroom environment".

When asked how principals might improve their supervisory and evaluative practices, 39 teachers responded. Two themes emerged and included *no recommendations* and *feedback*. First, overwhelmingly (n=21), teachers offered no suggestions regarding how principals could increase their supervisory and evaluative skills. Thirteen teachers believed principals could improve the quality of feedback they provided teachers, both positive and constructive. More specifically, teachers wanted more positive feedback about teaching strategies and desired constructive feedback about areas in which they were weak. For example, one teacher stated, "I would say teachers just need to be given more pats on the back throughout the year individually". Another respondent stated, "I've never had a bad evaluation. Come on! I can't be that awesome. I'd like to know where I need to improve."

Discussion and Conclusions

The purpose of this study was to understand teachers' perceptions about principals' supervision and evaluation in eight high performing elementary schools. The findings can be summarized into three general themes: (1) despite working in a school district with no differentiated teacher supervision and evaluation procedures, teachers' perceived principals engaged in all seven supervision and evaluation constructs including differentiated supervision and evaluation. However, differentiated supervision and evaluation was rated the lowest by all groups of respondents when compared to the other six supervision and evaluation constructs; (2) both non-tenured teachers and teachers with one to three years of experience were more positive about principals supervision and evaluation than more experienced teachers; and (3) teachers with five to 10 years of experience were the

least positive of any of the groups concerning principals' supervisory and evaluative behaviors.

Synthesizing specific findings, overall, teachers believed principals had adequate skills to effectively supervise and evaluate teachers. This confidence, along with principals' ability to form personal, meaningful relationships with teachers and treat them as professionals, creates a climate and culture that allows teachers to trust principals as supervisors of instruction. As a result, these principals understand trust is a precursor to effective teacher supervision (Zepeda, 2013). Teachers indicated principals routinely visited classrooms as part of their supervision responsibilities, a behavior Ikemoto et al. (2012) argued distinguished effective principals from ineffective principals and while encouraging teacher development. Teachers believed that because principals were in their classrooms weekly, principals understood the instructional climate of schools and positioned themselves to make instructional recommendations based on observable data (Ing, 2009). After classroom walkthroughs, teachers perceived positive feedback dispensed by principals as crucial for building morale and led to teachers' feelings of professionalism and shared leadership. However, teachers indicated they desired not only positive feedback, but also constructive criticism about their teaching so they could improve their performance.

Non-tenured teachers and teachers with one to three years of teaching experience were the most positive about principals' supervision and evaluation when compared to teachers with more experience. The researchers speculate non-tenured and novice teachers' positive views after their supervision hinge on two points. First, because principals engaged in a more direct approach when supervising non-tenured teachers and novice teachers (Glickman et al., 2005; Peterson, 1995; Nolan & Hoover, 2007; Zepeda, 2013), they possibly received more direct feedback about their performance. Second, because non-tenured and novice teachers struggle with technical teaching problems, principals can provide immediate feedback that helps and, as a result, non-tenured and novice teachers are more positive about the supervisory experience.

Results concerning the view of teachers with five to 10 years of experience are less clear as these teachers were the least positive about the seven supervision and evaluation constructs, and were significantly less positive about principals' ability to differentiate supervision and evaluation than early career teachers. These findings contradict others who describe this career stage of teaching as one in which teachers are mature, enthusiastic, and have high job satisfaction (Huberman, 1993; Zepeda, 2013). Attempting to describe why teachers with five to 10 years of experience were the most negative about principals' supervision and evaluation warrants further research. However, recent research indicates if experienced, effective teachers are provided no incentives because of school district policy restrictions, they tend to devalue teacher supervision and evaluation procedures (Jacob et al., 2012; Weisberg et al., 2009). Perhaps more importantly, in order to improve educational leadership training programs, researchers and practitioners should continue to work together to bridge the gap between theory and practice by identifying how teachers with five to 10 years of experience might be better supported with differentiated supervision and evaluation, allowing for more personalized professional development. Further research in this area

could help utilize the strengths of these veteran teachers and promote a greater component of shared leadership among staff members.

Regarding job assignment, support teachers rated six of the seven supervision and evaluation constructs higher than regular classroom teachers, indicating support teachers were more positive about principals' exhibiting these behaviors. Although these positive finding do not allow for conclusions of causality, two of the eight principals had backgrounds as content teachers. This might have assisted support teachers in expressing positive feeling about supervision and evaluation procedures in their schools, even within their content specific area.

In sum, the results of this study provide three conclusions. First, principals' instructional leadership is important for a positive instructional climate. Despite working in a school district with no differentiated supervision and evaluation processes, teachers in these eight schools were positive about all seven constructs concerning principals' supervision and evaluation, and believed principals' leadership led to feelings of trust and autonomy. Teachers reported principals were routinely in classrooms and viewed principals as instructional leaders.

Second, differentiated supervision and evaluation is difficult for principals as they attempt to provide assistance to all types of teachers. Even in these eight high performing schools, teachers perceived principals differentiated less than any other construct measured. Differentiated supervision and evaluation requires principals to have diagnostic skills when assessing the developmental level of teachers and match administrative supports to needs. Because differentiated supervision is not a *one size fits all* approach, it can be time consuming as principals collaboratively search out ways to remediate struggling teachers and enrich teachers meeting expectations. Compounding the difficulties in implementing differentiated supervision are school district evaluation procedures that treat all teachers the same. In such a context, principals have little autonomy to differentiate supervision and evaluation for teachers and are forced to navigate required paperwork requirements to meet policy demands. As a result, principals require professional development as a formal district process as they attempt to provide differentiated supervision for teachers. Thirdly, as teachers gain experience, their views of supervision and evaluation might become unfavorable, especially when compared to novice teachers. Principals must continue to challenge and support experienced teachers and attempt to keep their enthusiasm and desire for continuous improvement high. To do this, principals might provide frequent, public, positive feedback concerning effective teachers' performance and assist effective teachers in securing additional resources to further strengthen their instruction. Other examples include asking experienced, effective teachers to take a more active role in instructional leadership initiatives such as revising classroom walkthrough forms or helping with instructional analysis after classroom walkthrough data have been summarized. Finally, principals might allow experienced teachers to develop personalized professional development plans based on interests and worry less about requiring them to attend professional development not tailored to their innate desires. Thus, this study seeks to inform policies and practices for school districts looking to provide meaningful, ongoing professional development for their

employees. As teachers progress through various professional stages, school districts can better support the continual development of teachers by offering more developmentally appropriate supervision and evaluation that seeks to strengthen the organization as a whole.

References

Clark, P. L. (1998). *Perceptions of New Hampshire teachers and supervisors regarding teacher supervision*. (Unpublished doctoral dissertation). University of New Hampshire, Concord, NH.

Cuddapah, J. L., & Burtin, A. S. (2012). What all novices need. *Educational Leadership, 69*(8), 66-69.

Derrington, M. L. (2011). Changes in teacher evaluation: Implications for the principal's work. *The Delta Gamma Bulletin, 77*(3), 51-54.

Donaldson, M. L., & Donaldson, G. A. (2012). Strengthening teacher evaluation: What district leaders can do. *Educational Leadership, 69*(8), 78-82.

Jones, S. R., Torres, V., & Arminio, J. (2006). *Negotiating the complexities of qualitative research in higher education*. New York: Routledge.

Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2005). *The basic guide to supervision and instructional leadership*. Boston, MA: Pearson.

Gupton, S. L. (2010). *The instructional leadership toolbox: A handbook for improving practice*. Thousand Oaks, CA: Corwin.

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research*. Boulder, CO: National Teacher Education Policy Center. Retrieved from http://nepc.colorado.edu/files/PB-TEval-Hinchey_0.pdf

Huberman, M. (1993). *The lives of teachers*. New York: Teachers College Press.

Ing, M. (2009). Using informal classroom observations to improve instruction. *Journal of Educational Administration, 48*(3), 337-358.

Ikemoto, G., Taliaferro, L., & Adams, E. (2012). *Playmakers: How great principals build and lead great teams of teachers*. New York: New Leaders. Retrieved from http://www.newleaders.org/newsreports/publications/playmakers/

Jacob, A., Vidyarthi, E., & Carroll, K. (2012). *The irreplaceables: Understanding the real retention crisis in America's urban schools*. Brooklyn, NY: The New Teacher Project. Retrieved from http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf

Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership, 70*(3), 14-19.

Nolan, J. F., & Hoover, L. A. (2008). *Teacher supervision and evaluation* (2nd ed.). Hoeboken, NJ: John Wiley and Associates.

Peterson, K. D. (1995). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin.

Range, B. G., Scherz, S., Holt, C. R, & Young, S. (2011). Supervision and evaluation: The Wyoming perspective. *Educational Assessment, Evaluation, and Accountability, 23*(3), 243-265.

Range, B. G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers' think about their formative supervision in one US school district? *School Leadership and Management, 33*(1), 61-77.

Robertson, M. (2006). Why novice teachers leave. *Principal Leadership, 6*(8), 33-36.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). Brooklyn, NY: The New Teacher Project. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Zepeda, S. J. (2013). *The principal as instructional leader: A practical handbook* (3rd ed.). Larchmont, NY: Eye on Education.

Development, Validity, and Reliability of the Preservice Teachers' Attitude Toward Educational Research (P-TATER) Scale

Author Info:

Nathan E. Gonyea is an Associate Professor in the Department of Educational Psychology, Counseling, and Special Education at State University of New York College at Oneonta.

Correspondence regarding this manuscript should be addressed to Nathan E. Gonyea, Department of Educational Psychology, Counseling, and Special Education at State University of New York College, 108 Ravine Parkway, Oneonta, NY 13820. Ph: 607-436-3161. Email: Nathan.gonyea@oneonta.edu

ABSTRACT

While there is a growing emphasis on the use of data in educational decision making and research to select educational interventions, there is still limited research on the attitudes of teachers toward research and even less research on preservice teachers' attitudes. Additionally, there are few quantitative instruments designed to measure teachers' attitudes toward research and no instruments to measure preservice teachers' attitudes toward research. The instruments that do exist are of questionable quality. The purpose of the present study was to develop and measure of preservice teachers' attitudes called the Preservice Teachers Attitudes Toward Educational Research (P-TATER) scale.

Keywords: Attitudes, Educational Research, Preservice Teachers, Scale Development

Development, Validity, and Reliability of the Preservice Teachers' Attitude Toward Educational Research (P-TATER) Scale

Over the past several decades, researchers and policy makers have been increasingly emphasizing the use of research to inform K-12 teaching practices. At the same time there has been a growing awareness that teachers may not view education or educational research in the same way as researchers (Bulterman-Bos, 2008). Research investigating in-service and preservice teachers' attitudes toward educational research has not kept pace with this increasing awareness of the possible gap between researcher, policy maker, and teacher views. The research that has been conducted on teachers' attitudes tends to be sporadic and use inconsistent methods (Isakson & Ellsworth, 1979; Papasotiriou & Hannan, 2006; Short & Szabo, 1974).

Some studies have taken a qualitative approach, such as interviews (DeCorse, 1997; Papasotiriou & Hannan, 2006) and constructed response surveys (Galton, 2000; Hannan, Enright, & Ballard, 1998), while others have focused on quantitative scales (Eaker & Huffman, 1981; Isakson & Ellsworth, 1979; Short & Szabo, 1974). This lack of consistency makes it difficult to test the validity of these studies and impossible to gauge a change in teachers' attitudes over time. In order to address this gap in the research, valid and reliable instruments need to be developed to measure both preservice and in-service teachers' attitudes toward educational research across time, sample, and setting. The purpose of the present study was to develop a measure of preservice teachers' attitudes toward educational research in an effort to address this gap.

Review of Literature

Understanding teachers' attitudes toward research is becoming essential because of the need to train teachers in instructional techniques based on research such as response-to-intervention (RTI). Further, research on the impact of preservice teachers' beliefs indicates that these early beliefs have a significant impact on future teaching practice. Pajares (1992) conducted a comprehensive review of the literature on teachers' beliefs and found empirical support for the early formation of beliefs (i.e. prior to becoming teachers), the low likelihood of changing beliefs in adulthood, the persistence of beliefs even when confronted with strong scientific evidence that those beliefs are incorrect, and the significant impact of beliefs on teachers' perception, cognition, and behavior.

While our need to understand teachers' receptiveness to research based practices is increasing, our knowledge of teachers' attitudes toward research remains sparse and disjointed. The semi-structured qualitative interviews (DeCorse, 1997; Papasotiriou & Hannan, 2006) and constructed response surveys (Galton, 2000; Hannan, Enright, & Ballard, 1998) used in some studies can be informative, but lack generalizability. Adding to the difficulty of developing a comprehensive understanding of teachers' views of educational research is the small sample sizes of many of the most recent studies (DeCorse, 1997; Hannan, Enright, & Ballard, 1998; Papasotiriou & Hannan, 2006).

Quantitative scales can be generalized and allow for comparisons across samples and time, but only if they are valid, reliable, and used in multiple studies with diverse populations. However, in the past researchers developed new scales for each study, which were subsequently replaced by the next researcher with their own scale

making such comparisons impossible (Eaker & Huffman, 1981; Isakson & Ellsworth, 1979; Short & Szabo, 1974). Additionally, many of these surveys have been of questionable quality with little support for the validity of the individual items. For instance, Short and Szabo based their instrument on the 'research literature,' but never extrapolate on the type of literature or topics addressed by the literature. Isakson and Ellsworth and Eaker and Huffman developed their scales based on their personal knowledge/experiences.

A slight deviation from this trend was the scale developed by Ozturk (2011), who modified Isakson and Ellsworth's (1979) scale. However, Ozturk's method had several limitations. First, Oztruk's measure was developed by conducting an exploratory factor analysis on the measure developed by Isakson and Ellsworth. While this may appear to support the validity of the final measure, the lack of rigorous procedures in the development of the original instrument undermines the validity of the final instrument. Isakson and Ellsworth based their original instrument on their personal knowledge and experiences and not on any actual data on teachers' attitudes. In order for researchers to develop a comprehensive knowledge of teachers' views of research, an instrument grounded in these views needs to be developed. Additionally, an instrument needs to be developed that addresses the attitudes of teachers earlier in their careers. The present study develops an instrument based on a qualitative survey of preservice teachers' attitudes toward research.

Research Questions

In order to develop the instrument the following research questions were posed for each stage of the study. Stage 1: What are the major themes in preservice teachers' attitudes toward research? Stage 2: What is the underlying structure (i.e. factors) of preservice teachers' attitudes toward research? Which items, based on the stage 1 thematic analysis, load most highly on each factor? Is the new scale valid and reliable? Stage 3: Can the validity and reliability of the *Preservice Teachers' Attitude Toward Educational Research Scale* be supported using a different sample?

Method

Design

The present study used a 3 stage design. Stage 1 involved administering an 8-item constructed response survey to undergraduate education majors. A thematic analysis of responses was conducted using a consensual qualitative approach as detailed by Barbara, Thompson, and Williams (1997). The coding team consisted of 1 education faculty member and 3 undergraduate education majors. All coding was audited by a different education faculty member to ensure the validity of the codes. In stage 2, representative statements of students' responses to the stage 1 survey were constructed based on the identified themes. These statements were used to construct a 49 item Likert-type questionnaire, which was administered to a sample of education majors, and subjected to an exploratory factory analysis. The factor structure identified in stage 2 was evaluated in stage 3 using a second exploratory factor analysis and a separate sample of undergraduate education majors.

Participants

All three stages used samples of preservice undergraduate education students. The sample used in stage 1 consisted of 46 juniors (n = 24) and seniors (n = 22) enrolled in core education classes. The majority were female (n = 36) and they had

a median age of 21 years. The stage 2 sample consisted of 215 sophomore, junior, and senior students enrolled in core education classes. The participants were predominately Caucasian (n = 203), female (n = 167), elementary education majors (n = 138), with a mean age of 20 years. The sample in the third stage included 255 sophomore, junior, and senior students enrolled in core education classes. They were primarily Caucasian (n = 240) and female (n = 192). There were approximately equal numbers of elementary (n = 138) and secondary (n = 115) education majors. Participants had a mean age of 21. All education students at the college are exposed to a discussion of the connection between educational research and teaching practice in their introductory level education courses. Over 90% of students in all three samples reported remembering discussing research in their education courses.

Instruments

Three questionnaires were used in the present study. The stage 1 questionnaire consisted of 8 constructed response items (See Tables 1-3). These items addressed preservice teachers' attitudes concerning the usefulness of educational research, their competence in understanding and conducting educational research, and the role teachers should play in educational research. Students were first asked to answer yes or no to each question and then provide a written explanation of their answer. The stage 2 questionnaire consisted of demographic questions and 49 Likert-type items developed from the themes identified in the thematic analysis of the stage 1 questionnaire (See Table 4). The Likert-type items were rated on a 6-point scale ranging from strongly agree to strongly disagree. For stage 3, the questionnaire consisted of demographic questions and the 33 Likert-type items that were selected for inclusion in the final scale in stage 2 (See Table 4).

Results

The responses to the stage 1 questionnaire were analyzed using a thematic analysis and a total of 44 themes were identified. Between 4 and 6 themes were identified for each question. The 8 questions, associated themes, and the number of comments that addressed each theme can be found in tables 1-3. Most themes aligned with comments from students who answered yes or no. However, 3 of the themes include comments from people who answered yes and no. These 44 items were used to generate 49 statements, which were the basis for the *Preservice Teachers Attitudes' Toward Educational Research (P-TATER) Scale*. All themes were used to create statements, including those reported by a small number of participants in order to ensure that the final list of questions addressed all possible aspects of preservice teachers' attitudes toward research.

Insert Tables 1-3 About Here

The 49 items developed based on the stage 1 results were analyzed using an exploratory factor analysis in stage 2. The principal factor method was utilized. Costello and Osborne (2005) propose that a true factor analysis method is better than principal components analysis because true factor analysis methods separate shared and unique variance, principal components analysis does not. The principal factor extraction method was chosen over other extraction methods because of the non-normal distribution of some of the items. Some authors indicated that the principal factor method of extraction is a more robust method when dealing with non-normality of variables (Costello and Osborn, 2005;

Fabrigar, Wegener, MacCallum, & Strahan, 1999). The factor analysis was first run without rotation and then with an oblimin rotation. However, there were no significant differences in factor loadings, so only the unrotated results are reported. The Catell scree test was used to identify the number of underlying factors (Cattell, 1966) and using this method two underlying factors were identified.

The factor correlation matrix shows that 33 of the items had factor loadings of .40 or above on one of the 2 factors (See Table 4). Any item with a factor loading of .40 or above was included in the factor. Two items (43 & 44) had factor loadings of .40 or above for both factors. These items were included in the scale for which they had the higher loading; the attitude scale. A factor loading of .40 was chosen based on de Winter, Dodou, and Wieringa's (2009) estimates of necessary sample sizes and factor loadings for satisfactory factor recovery. Using de Winter et al.'s estimates for analyses with 2 factors, at least 24 variables, and minimum factor loadings of .40 the minimum estimated samples size is 134.

The first factor, the attitude scale, consisted of 25 questions relating to perceptions of the value and usefulness of educational research for teachers. The second factor, the perceived competence scale, consisted of 8 questions relating to the preservice teachers perceptions of their ability to understand and conduct research.

Insert Table 4 About Here

Both scales are cumulative, with each item scored from 1 (Strongly Disagree) to 6 (Strongly Agree). Items with negative factor loadings were reverse scored. The attitude scale has 5 inverse items, while the perceived competence scale has 1. A reliability analysis of both scales using coefficient alpha indicated high reliability for both the attitude scale ($\alpha = .93$) and the perceived competence scale ($\alpha = .82$).

In stage 3, the 33 items selected for inclusion in the *P-TATER* were analyzed using a second exploratory factor analysis. The principal factor method of extraction was used again without rotation with a 2 factor structure specified. All of the items loaded most strongly on their original factor. All but one of the factor loadings were above .4 (See Table 5). Similarly, the alpha reliabilities of the attitude and perceived competence scales were higher than in stage 2, $\alpha = .94$ and $\alpha = .85$ respectively. The two factors account for 41% of the variance in the items' variance-covariance matrix.

Insert Table 5 About Here

Discussion

The data indicate that the *Preservice Teachers' Attitude Toward Research (P-TATER) Scale* has potential as a measure of preservice teachers' perceptions of research. In contrast to previous measures (Eaker & Huffman, 1981; Isakson & Ellsworth, 1979; Short & Szabo, 1974), the *P-TATER's* items are grounded in preservice teachers' comments about research. This inductive development of items, in conjunction with the high factor loadings and conceptually consistent grouping of items into the 2 scales, supports the validity of the *P-TATER* as a measure of teachers' attitudes toward research. The Cronbach's alpha reliability coefficients for *P-TATER* are similar to those of Isakson & Ellsworth (1979) and greater than those reported by Ozturk (2011), the only other two studies to report reliability coefficients.

Given the lack of agreement on how to determine appropriate minimum sample size (de Winter et al, 2009), the relatively small size of the present sample is a limitation. The present study applied the minimum sample size estimate identified by de Winter et al (2009) through their Monte Carlo analysis using varying factor loadings, number of factors, and number of variables. The present study's samples were greater than the minimum identified by de Winter et al. given the present study's parameters. However, the sample sizes were below the rule of thumb participant per item estimates given by other authors (Schreiber, Stage, King, Nora, & Barlow, 2006). Future research should use larger sample sizes in line with participants per item estimates to confirm the stability of the factor structure.

The composition of all three samples used in this study were consistent with the demographics of the undergraduate education students at the college used in this study and the current composition of P-12 teachers in the United States (i.e. predominately female and Caucasian) (United States Department of Education, National Center for Educational Statistics, 2010). Despite this being the case, the instrument should be evaluated with a more diverse sample to determine whether or not it can be used with all preservice teachers. Further, the factor structures' stability between different types of education students (e.g. elementary vs. secondary or science education vs. English education) should also be evaluated.

In order to avoid past cycles of continual redevelopment of teacher attitude scales, researchers need to conduct further analyses of the P-TATER's validity and reliability, as well as its generalizability across time and population. Additionally, the efficacy of using the P-TATER, or a derivative of the P-TATER, to evaluate in-service teachers' views of educational research should be investigated. To further these efforts, the P-TATER should be made available for broad evaluation by educational researchers and use by teacher education faculty.

The two subscales (Attitudes and Perceived Competence) can provide useful information for teacher education faculty. These two subscales are particularly relevant given Pajares (1992) conclusions concerning the low likelihood of changing beliefs in adulthood, the persistence of beliefs even when confronted with strong scientific evidence that those beliefs are incorrect, and the significant impact of beliefs on teachers' perceptions, cognition, and behavior. Based on Pajares' conclusions, students' general attitudes toward the usefulness and importance of educational research (Attitudes scale) and competence (Perceived Competence scale) are likely to predict students' likelihood of evaluating teaching methods based on research evidence and adopting those methods. Further, Pajares' conclusions would suggest that it is unlikely for students to change their attitudes based on "evidence" provided in class, and that education faculty would need to convince students who have a low score on the Attitude Scale that the gestalt of the education community supports the use of empirically based methods in order to create a change in belief.

References

Bultermane-Bos, J. A. (2008). Will a clinical approach make education research more relevant for practice? *Educational Researcher, 37*(7), 412-420. doi:10.3102/0013189X08325555

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. doi:10.1207/s15327906mbr0102_10

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7). Retrieved from: http://pareonline.net/pdf/v10n7.pdf

DeCorse, C. J. B. (1997, March). *"I'm a Good Teacher, Therefore I'm a Good Researcher": Changing Perceptions of Expert and Novice Teaches about Doing Research*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED411-267). Retrieved from http://files.eric.ed.gov/fulltext/ED411267.pdf

Eaker, R. E., & Huffman, J. O. (1981). *Teacher Perceptions of Dissemination of Research on Teaching Findings*. (Occasional Paper No. 41). East Lansing, MI: The Institute for Research on Teaching (ERIC Document Reproduction Service No. ED205-501). Retrieved from http://eric.ed.gov/?id=ED205501

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, R. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. doi:10.1037/1082-989X.4.3.272

Galton, M. (2000). *Integrating Theory and Practice: Teachers' Perspectives on Educational Research.* Retrieved from www.tlrp.org/pub/acadpub/galton2000.pdf

Hannan, A., Enright, H., & Gallard, P. (1998). Using research: The results of a pilot study comparing teachers, general practitioners, and surgeons. *British Education Index - Education-line*. Retrived from www.leeds.ac.uk/educol/documents/000000851.htm

Isakson, R. L., & Ellsworth, R. (1979). The measurement of teacher attitudes toward educational research. *Educational Research Quarterly, 4*(2), 12-18.

Ozturk, M. A. (2011). Confirmatory factory analysis of the educators' attitudes toward educational research scale. *Educational Science: Theory & Practice, 11*(2), 737-747.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Clearning up a messy construct. *Review of Educational Research, 62*(3), 307-332. doi:10.3102/00346543062003307

Papasotiriou, C., & Hannan, A. (2006). The impact of education research on teaching: the perceptions of Greek primary school teachers. *Teacher Development, 10*(3), 361-377. doi:10.1080/13664530600922328

Schreiber, J. B., Stage, F. K., King, J., Nora, A., Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-337. doi:10.3200/JOER.99.6.323-338

Short, B. G., & Szabo, M. (1974). Secondary school teachers' knowledge of and attitudes toward educational research. *The Journal of Experimental Education, 43*(1), 75-78.

U.S. Department of Education, National Center for Education Statistics. (2010). Teacher attrition and mobility: Results from the 2008–09 teacher follow-up survey (NCES Publication No. 2010-353). Retrieved from http://nces.ed.gov/pubs2010/2010353.pdf

Table 1

Do preservice teachers perceive themselves as having the skills necessary to understand and apply research?

	Yes	No
Have you discussed research in your teacher preparation course work?		
brief class mention	11	0
class discussion	11	0
did research/ used research for project or paper	7	0
example/ support lecture	4	0
no discussion	0	13
Do you feel confident in your ability to read and understand educational research articles?		
confusing jargon/ stats	0	9
if you read/ analyze carefully	7	0
lack of practice	0	4
previous experience/ learn in class	13	0
depends on ease of article	3	0
want more confidence	0	5
Do you feel confident in your ability to apply information you read in educational research articles to your teaching?		
amount of time required	4	1
lack of experience/ability	0	14
need to be more comfortable	0	2

research is to theoretical/ Impractical	0	2
Relevance	1	1
prior knowledge/ ability/ practice	11	0

Table 2

What role do preservice teachers believe educational research has in their classroom?

	Yes	No
Should teachers stay up-to-date on current educational research findings?		
avoid misinformation	2	0
changing research	7	0
continuous learning	3	0
improve teaching/ classroom practice	22	0
stay informed	19	0
Should teachers modify their teaching practices based on recent education research findings?		
improvement/ continuous learning	5	0
effective teaching	15	0
valid & relevant	23	0
some old methods work	0	9
Is educational research useful for classroom teachers?		
keep up to date	7	0
only use valid research	4	0
provide solutions	2	0
test current methods	4	0
improve teaching/ classroom practice	3	0
useful in class/ to students	11	0

Table 3

Do preservice teachers believe that classroom teachers should be involved in conducting research?

	Yes	No
Should teachers be involved in conducting educational research?		
Biased	0	2
if they have time	7	0
if interested	8	0
experience/ expertise	16	0
lack of time	0	5
give teachers useful information	13	0
Should teacher preparation programs include a course on conducting educational research?		
learn beneficial skills	23	0
other things are more important	0	5
identify good research	3	0
Optional	1	3
graduate programs only	0	2
not a separate class	0	6

Table 4

Stage 2 Factor Analysis of Initial P-TATER Questions

Item (Item number on final P-TATER scale)[mean(standard deviation)]	A	P. C.
1. It is important for teachers to stay up-to-date on educational research in order to stay informed of new findings related to teaching and learning. (1)* [5.6(.56)]	.53	.05
2. Teachers shouldn't be involved in conducting educational research because their professional practice has made them biased. [2.75(1.14)]	-.23	.08
3. Reading educational research is a waste of a teacher's time. (2)* [1.64(.65)]	-.59	.05
4. Teachers should only be involved in conducting educational research if it personally interests them. [2.76(1.26)]	-.27	-.12
5. Teachers should stay up-to-date on current educational research in order to avoid misinformation about teaching and learning. (3)* [5.28(.71)]	.59	.04
6. Conducting research distracts teachers from other more important activities. (4)* [2.57(1.06)]	-.41	.07
7. Teachers can learn more about new teaching practices from in-service trainings, seminars, and conferences than they can by reading educational research. [4.53(1.16)]	-.02	.01
8. Teachers don't have enough time to be involved in conducting research. [2.80(1.04)]	-.38	-.06
9. Teachers should read current educational research because it promotes continuous learning. (5)* [5.29(.72)]	.65	.06

10. Teachers' firsthand experience in the classroom is a valuable asset to educational research studies. [5.36(.77)]	.33	-.01
11. Reading educational research can help teachers improve their teaching. (6)* [5.21 (.75)]	.70	-.07
12. Conducting educational research provides teachers with valuable information/experiences that will help improve their classroom practices. (7)* [5.04(.79)]	.68	-.07
13. Teachers should *not* change teaching practices because some old methods may be more effective. [3.19(1.19)]	-.24	.03
14. Educational research classes can teach preservice teachers valuable skills. (8)* [4.96(.72)]	.60	.12
15. Teachers should change their teaching based on research that investigates students similar to their own. [3.97(.94)]	.31	.01
16. Educational research classes should be optional for education students.[3.66(1.29)]	-.32	.01
17. Teachers should make changes to their teaching based on action research. [4.12(.92)]	.36	-.06
18. Educational research instruction should be imbedded in other education classes. (9)* [4.46(.94)]	.60	-.12
19. Practical experience is a better guide to effective teaching methods than research. [4.48(1.01)]	-.34	.18
20. Teacher preparation programs should include an educational research class. (10)* [4.32(1.02)]	.58	-.02
21. Teachers should *not* attempt to apply theoretical research to their teaching. [2.92(1.05)]	-.32	.03
22. Educational research courses should be required for education graduate students. (11)* [4.23(1.16)]	.51	-.17
23. Teachers should make changes in their classroom based on applied educational research. (12)* [4.30(.82)]	.49	-.03
24. Undergraduate teacher education programs should *not* include a course on conducting educational research. (13)*	-.53	.07

[3.11(1.15)]		
25. Teachers should modify their teaching based on recent educational research because they need to continue to improve their practice. (14)* [4.39(.95)]	.54	.08
26. Teachers should *only* modify their teaching based on current educational research if they believe their current methods are not working. [3.95(1.23)]	-.27	.03
27. There are classes that are more important than research methods in teacher preparation programs. [4.73(1.05)]	-.26	.11
28. Educational research helps teachers stay up-to-date on effective teaching methods. (15)* [5.05(.75)]	.73	.01
29. I do not have enough practice reading and interpreting research articles to do it effectively. (16)** [3.91(1.26)]	.13	-.57
30. Educational research is too theoretical to be useful. (17)* [2.80(.96)]	-.55	-.07
31. Teachers can gain new knowledge from educational research. (18)* [5.26(.61)]	.67	.00
32. I feel confident in my ability to apply action research findings in my classroom. (19)** [4.09(1.07)]	.12	.69
33. Educational research helps teachers improve their practice. (20)* [4.85(.75)]	.76	.07
34. Any teacher can understand research if they read it carefully. [3.71(1.18)]	.11	.38
35. I can successfully apply research I read to my teaching. (21)** [4.20 (.90)]	.17	.70
36. Educational research can help teachers meet their students' needs. (22)* [4.75(.82)]	.63	.17
37. Research articles are straight forward and easy to understand. (23)** [2.57(.95)]	.03	.40

38. I feel confident in my ability to apply theoretical research findings in my classroom. (24)** [3.73(.96)]	.20	.79
39. Educational research can provide solutions to classroom problems. (25)* [4.57(.80)]	.49	.17
40. I would need to take a class on educational research before I would feel confident reading educational research articles. [4.08(1.28)]	.20	-.38
41. I feel confident in my ability to use applied research findings to modify my instruction. (26)** [3.90(.93)]	.10	.72
42. Research can help teachers by testing methods teachers use in the classroom. (27)* [4.81(.71)]	.54	.09
43. I would like to become more confident in my ability to read educational research. (28)* [4.76(1.09)]	.54	-.41
44. I need to be more experienced in applying research to my classroom. (29)* [4.79(1.01)]	.46	-.41
45. Summaries that interpret and integrate research are more useful than research articles for teachers. [4.11(1.03)]	-.11	-.06
46. I am confident in my ability to understand research jargon. (30)** [3.11(1.11)]	.07	.52
47. I have lots of experience applying research findings to my teaching. (31)** [2.34(1.07)]	-.04	.52
48. Students benefit when their teachers are up-to-date on current educational research. (32)* [4.99(.89)]	.65	-.01
49. I *do not* plan on applying research findings in my classroom. (33)* [2.20(.96)]	-.54	.00

A = Attitude

P.C. = Perceived Competence

* Item included in the final Attitude Scale

** Item included in the final Perceived Competence Scale

Table 5

Stage 3 Exploratory Factor Analysis of Items for the P-TATER Scale

	Attitudes	Perceived Competence
1.*	.58	-.04
2*†	.57	-.04
3*	.61	.03
4*†	.48	-.07
5*	.72	.00
6*	.70	.07
7*	.72	.00
8*	.64	-.09
9*	.60	.01
10*	.66	-.10
11*	.59	-.16
12*	.59	-.11
13*†	.62	-.17
14*	.60	-.12
15*	.68	-.06
16**†	.13	.47
17*†	.50	-.08
18*	.65	-.15
19**	.36	.70

20*	.75	-.09
21**	.31	.62
22*	.72	-.09
23**	.21	.35
24**	.35	.74
25*	.59	-.07
26**	.32	.79
27*	.59	-.06
28*	.50	-.29
29*	.49	-.38
30**	.23	.55
31**	.09	.53
32*	.62	-.05
33*†	.63	-.07

* Included in the Attitude Scale

** Included in the Perceived Competence Scale

† Reverse Coded Item (Items were reverse coded before factor analysis)

Student Response to Faculty Instruction (SRFI): An Empirically Derived Instrument to Measure Student Evaluations of Teaching

Author Info:

Brian D. Beitzel, Department of Educational Psychology, Counseling and Special Education, SUNY Oneonta.

Correspondence concerning this article should be addressed to Brian D. Beitzel, Department of Educational Psychology, Counseling and Special Education, SUNY Oneonta, 20 Denison Hall, Oneonta, NY 13820. E-mail: brian.beitzel@oneonta.edu

Abstract

The Student Response to Faculty Instruction (SRFI) is an instrument designed to measure the student perspective on courses in higher education. The SRFI was derived from decades of empirical studies of student evaluations of teaching. This article describes the development of the SRFI and its psychometric attributes demonstrated in two pilot field tests. Students in all levels of courses (freshman through graduate) and from many different class sizes participated in these two pilot studies. Response rates were high and the ratings students assigned to the SRFI items corresponded with course characteristics in ways consistent with the empirical literature. Through these two pilot studies the SRFI has shown itself to be a reliable and valid instrument for ascertaining the student perspective on instruction in higher education.

Keywords: student evaluations of teaching, reliability, validity

Student Response to Faculty Instruction (SRFI):

An Empirically Derived Instrument to Measure Student Evaluations of Teaching

Faculty in higher education need no introduction to student evaluations of teaching (SETs). Privately and publicly, these instruments are often maligned as a special kind of curse on faculty who have contributed their earnest energies in a sincere effort to be the best instructors they can be. When the end of the term arrives, students are invited to voice their opinions about the course and the instructor—and almost invariably, one or more students (particularly those who lacked personal responsibility for their performance in the course) will not see the instructor in the most favorable light. At many institutions, data from these instruments weigh heavily in tenure and promotion decisions (Baldwin & Blattner, 2003; Lindahl & Unger, 2010). The perception among many faculty is that they could be punished (i.e., with a negative personnel decision) for something they did not do.

A portion of this distaste is justified. Data from SETs are regularly reviewed by peers (e.g., on tenure and promotion committees) and administrators, few of whom have any training in the proper interpretation of these data (McKeachie, 2007). Wilbert McKeachie, a well known and highly respected scholar of college teaching, expresses his view of this situation unequivocally: "The major validity problem is in the use of the ratings by personnel committees and administrators" (McKeachie, 1997, p. 1222). Only a handful of dossier reviewers possess the discernment to recognize what the instrument was designed to measure, what its limitations are, and what evidence exists to demonstrate its validity. When confronted with a veritable mountain of data containing numbers to the hundredths decimal place, one could feel compelled by this ostensible precision to arrive at a conclusion (positive or negative) that may be unwarranted, perhaps without even being aware that such definitive-looking numbers do not represent the certainty we wish they would. This perception of accuracy can result in unjustifiable actions, such as ranking faculty for merit considerations on the basis of SET ratings. According to McKeachie, "attempting to compare teachers with one another by using numerical means or medians" is a regrettable use of SET data (1997, p. 1222).

Worse yet, SETs are not always developed by measurement professionals. In these cases, faculty are quite justified for their concerns about the integrity of the instrument. On my campus, a SET was developed and used for 20 years without any formal examination of its validity. Several of the 11 items on this instrument are of suspicious worth—for example, "Classes met regularly as scheduled." It turns out that those of us who have met every class as scheduled for the entire semester still do not get perfect scores on this item. Thus, because the data for this item are demonstrably inaccurate, the item is of dubious value. (One may be allowed to wonder what other items may be garnering responses that are similarly inaccurate.) Likewise, the item "Methods of assessment were graded fairly" poses a problem for instructors who hold less-than-ambitious students to high standards; many times these students seem to believe that grades which are "fair" are those in the "A" range.

There is one further point to make about instruments that are developed by non-experts. Many SETs ask students to respond to a final item that says something like, "Overall, this is a good instructor." There is

nothing at all wrong with this kind of "global" item. However, there is an extensive cognitive-science literature demonstrating indisputably that prior-knowledge activation (or "priming") has an effect on subsequent responses (e.g., Bransford & Johnson, 1972; Kinchla, 1992; Logan, 1980; Masson, 1995; McNamara, 1992; Ratcliff & McKoon, 1995; Schacter & Cooper, 1995). In the case of SETs, this means that whatever ideas students have been asked to think about in earlier items could influence how they respond to later items. Specifically, asking students to think about whether classes met as scheduled and whether they received the grade they thought was fair "primes" them for the final global item about the overall quality of the instructor. Regardless of the rigor of the course or the practical value of the required activities (ideas which are not activated by questions on this instrument), students' opinions about the overall quality of teaching are influenced by the ideas most recently activated. After being prompted to remember, among other things, that the teacher held class every day and gave high grades for all student work, the trajectory leads toward marking that person as a good teacher overall. Now it should be easy to see how one could legitimately question the data for even a global item if it follows other items asking about teaching behaviors which even the most ineffective of instructors can easily produce. I have never seen an empirical test of this in the SET literature, but it would be an excellent (perhaps shocking) study.

Because SETs have become necessary in higher education, we must strive to advocate for the highest quality instruments we can achieve within our inherently political environments. SETs that are generated exclusively through group discussion and idea-sharing—without access to empirical data or measurement expertise—are unequivocally inadequate. Higher education, of all places, ought to pursue such endeavors in the most informed manner possible.

Development of the SRFI

The purpose of this article is to describe the development and psychometric properties of the Student Response to Faculty Instruction (SRFI, pronounced *SIR-fee*), a new SET derived from many empirical studies of SETs conducted over the last several decades. Because of the above-described problems with our existing instrument, we (the Committee on Instruction, of which I was chair) elected to abandon the existing instrument and develop a new one.

Purpose of the SRFI

For practical reasons, the SRFI was designed to be an instrument providing summative feedback about a course, from the students' perspective, to both the instructor and administration (chair, dean, etc.). Such information could also be used formatively for future course improvement; however, the principal goal was to obtain high quality summative data.

Guiding Principles

At the outset, our chief goal was to develop an instrument that would be general enough to cover all types of courses at our institution (e.g., freshman, senior, graduate, large, small, general education, distance-learning) to improve the concurrent situation in which some faculty felt they were forced to administer an instrument that was not fully compatible with their course.

A second goal was to have as empirical a rationale as possible for inclusion and exclusion of any potential item to minimize the disagreements that would inevitably result from faculty advancing their own

ideas of what should be included or excluded.

To help us meet this second goal, we settled on Feldman's (2007) categorization scheme as our guiding framework. Essentially, he uses two empirical mechanisms to determine which dimensions of teaching are the most important for identifying effective instruction. First, some specific qualities of an effective teacher should correlate highly with student achievement. On average, the courses with the highest performing students are likely taught by instructors with at least some of these superior qualities (Cashin & Downey, 1992). For SETs, this translates to looking for the strongest positive correlations between student achievement (e.g., final exam scores) and specific SET items. Thus, the items that have the strongest positive correlations with student achievement are expected to be the best items for identifying effective teachers. Such items provide us with the best opportunity to escape the priming effect that I described earlier with regard to global items. In Feldman's analysis, items measuring the "Clarity and Understandableness" of the teacher had one of the strongest positive correlations with student achievement. In contrast, items measuring the "Nature and Usefulness of Supplementary Materials and Teaching Aids" had a negative correlation with student achievement.

Feldman's second mechanism for empirically discriminating teacher quality is student responses to global items (e.g., "This is an excellent instructor"). The rationale for this is that if a teacher is an excellent instructor overall, that person must surely also be rated highly on some of the more specific (non-global) SET items. The SET items that meet this second mechanism are those that correlate highly with one or more global items. For example, Feldman found that items addressing the "Teacher's Availability and Helpfulness" correlated most highly with global items. Conversely, items measuring the "Teacher's Stimulation of Interest in the Course and Its Subject Matter" had the weakest correlation with global items.

When selecting potential items for a new instrument, one could rather easily be justified by including an item when it complies with either one of Feldman's mechanisms. However, we specified that *both* mechanisms must be operative in order for us to include an item on our new instrument. That is, in order to be considered for inclusion on our new instrument, a potential item must conceptually fit within one of Feldman's categories having the strongest positive correlations with (a) student achievement *and* (b) global ratings. Because Feldman's mechanisms were founded upon decades of empirical work, we had a solid empirical basis for item selection.

Process

After our selection criteria were established, we generated a pool of items that followed our guiding principles described above. These items were derived from examining other similar instruments, feedback from our campus colleagues via an open-ended survey, Feldman (2007) and other published literature (e.g., Marsh, 1982a; Spooren, Mortelmans, & Denekens, 2007), and our existing campus instrument. The SEEQ (Students' Evaluations of Educational Quality; Marsh, 1982a) is a well known and highly respected instrument for ascertaining student opinion of instruction in college classes. Spooren et al. (2007) developed an instrument that has both a theoretical and empirical basis, but was piloted on a limited set of courses; nonetheless, their sub-dimensions were useful in constructing an

initial item pool consistent with Feldman's (2007) framework. We modified the items as necessary for clarity and to avoid copyright infringement.
We did not seek to attain coverage of all of Feldman's categories. Rather, we sought to design an instrument that was in the range of 7-12 items that met the above-mentioned criteria. After assembling an initial item pool, we subjected our fledgling instrument to a pilot test.

Pilot Study 1
Method

Through my committee's informal network of colleagues around campus, we recruited 42 classes of diverse types (large, small, freshman, upper-division, graduate) to participate in the first pilot study. Of the classes that participated, 7 were 100-level, 20 were 200-level, 10 were 300-level, and 5 were graduate classes. The median enrollment for these classes was 20 students. A third party (a secretary) was the only person who maintained and had access to the participant list. All communication between me and the participating faculty was directed through her to maintain confidentiality.

Pilot Instrument 1

The instrument for Pilot Study 1 had 22 items to be rated on a 5-point scale ("Strongly Disagree," "Disagree," "Neutral," "Agree," and "Strongly Agree"). Many of these items were being tested for inclusion on the final instrument; these are referred to throughout this article as "pilot items"; the subset of pilot items that were eventually included on the final version of the SRFI are referred to in this article as "SRFI items."

Validation Items.

In addition to the pilot items, several other items were included on the pilot instrument for validation purposes, described in the succeeding paragraphs. These items are referred to throughout this article as "validation items."

Student Motivation.

One of the strongest correlates with SET ratings is students' prior interest in the subject (Marsh, 2007). The more motivated a student is in a specific course, the higher that student's grade is likely to be—and the more the student should rate the instructor favorably (Cashin, 1995). To test this assumption, we included three motivational items on our pilot instrument to look for relationships between them and the pilot items.

Quality of Instructional Goals.

Similarly, we included one item to determine whether the instructors who had course goals that were higher than simply memorization of the material would receive different ratings than courses in which students felt their principal task was memorization. In the SET literature, courses that are more difficult tend to receive higher student ratings (with the exception of extremely difficult courses; Marsh, 2007).

Knowledge Gain.

Students who believe they have learned a lot in the course are more inclined to rate a professor favorably than students who do not perceive they have learned much in the course (Cashin, 1995; Feldman, 2007; Stapleton & Murkison, 2001). One item related to perceived knowledge gain was included on the pilot instrument to test for this relationship.

Global Items.

We included three global validation items (beginning with the word "Overall") to assess the relationship between these global items and the pilot items (Feldman's second mechanism for determining teacher quality from student ratings, discussed earlier). These global items were presented as the last items on the pilot instrument.

Results and Discussion

Response Rate.

Overall, 775 students (81%) responded to the pilot instrument. By practically any standard, this is a very satisfactory response rate. Among the 37 undergraduate classes, 738 students (80%) responded, and from the 5 graduate classes 37 students (95%) responded.

Item Reduction.

One goal of Pilot Study 1 was to narrow down the pool of items to a final size of 7-12 items (see the appendix for the final set of items). To determine which items to keep, a stepwise regression analysis was conducted, using a criterion of $p = 0.05$ for entry and $p = 0.10$ for exit. The dependent variable was each of three global validation items. Thus, three regression analyses were conducted. In order to be included on the final version of the SRFI, we determined that a pilot item must have been a predictor in *at least two* of the three regression analyses in *both* Pilot Study 1 *and* (eventually) Pilot Study 2. These stringent criteria were designed to help us select only the highest quality, most reliable items for the final instrument. Multicollinearity is a concern for analyses of this type, due to the expected correlation among typical SET items. To examine this assumption, the coefficients, their standard errors, and the significance tests were visually inspected at each step of these stepwise regression models. Across the entire series of regressions reported in this article (in both Pilot Study 1 and Pilot Study 2), the coefficients and their standard errors were remarkably stable, and there were no cases in which the significance tests were inconsistent across each progressive step of the regression (cf. Kidwell & Brown, 1982). Thus, we can state with confidence that the independence assumption associated with the multiple-regression analyses has been met.

Regression Results.

All regression models had omnibus *p*s less than 0.0001. The final (stepwise) models contained between 5 and 8 predictor variables, demonstrating that we do not need a very large number of items to make a satisfactory prediction about students' views regarding their instructors' overall teaching skill. The final models included at most one item that was not eventually included on the final version of the SRFI; the non-included item was not the same item in all three models. Thus, we did not eliminate any item that the regression analyses uniformly indicated as highly useful in predicting responses to the global validation items. The R^2 values for these final stepwise models ranged from 0.707 to 0.841; therefore, we can account for a modestly strong proportion of the variance in the global validation items with a small subset of the pilot items. The pilot items that entered the regression models but were not included on the final version of the SRFI entered no earlier than Step 4 of the regressions. Thus, we did not eliminate any items that were most predictive of overall teaching quality.

Pilot Items Eliminated. Several pilot items were eliminated on the basis of the regression results because they were not predictors in more than one of the three regression models. For example, the three

items, "The instructor was prepared for class," "The course was well organized," and "The instructor responded satisfactorily to questions both in and out of class" were predictors for exactly *none* of the three global items and were therefore eliminated from further consideration.

Global-Like SRFI Item. The single global-like item retained on the final version of the SRFI was, "I would recommend this instructor to other students." This item was the first to enter in each of the three stepwise regressions, with *F*-ratios ranging from 670.09 to 2697.83, and R^2 values ranging from 0.60 to 0.78. Therefore, this item alone captures most of the variance in the global validation items. In fact, one would not be poorly served by asking this question alone!

Reliability.

Reliability was calculated in three different ways. The well known, widely used, and often-criticized (cf. Cortina, 1993; Green & Yang, 2009; Sijtsma, 2009) coefficient alpha (Cronbach, 1951) yielded an internal consistency measure of 0.93.

Following the advice of Revelle and Zinbarg (2009), the ω_t statistic was also calculated using the `psych` package from the open-source software R (R Development Core Team, 2011). The result was also 0.93.

Marsh (1982b) recommended that an intraclass correlation is the most appropriate statistic for estimating the reliability of SETs because it anticipates high levels of agreement within each course and different ratings among courses. For Pilot Study 1, the intraclass correlation was 0.93.

Thus, from three different approaches to measuring reliability, the results are all 0.93. For the pilot version of any instrument, this level of reliability is a measurement specialist's dream.

Uni-Dimensionality.

There has been some interest within the SET literature regarding the dimensionality of SET instruments (Abrami, d'Apollonia, & Rosenfield, 2007; d'Apollonia & Abrami, 1997). To test the dimensionality of the SRFI, the Pilot Study 1 data were randomly split into halves, and an exploratory factor analysis was conducted, using only the SRFI items (after the other pilot items had been eliminated) and extracting only factors with eigenvalues greater than 1. Using principal components extraction, only one component was derived, with loadings all greater than 0.72. These high loadings indicate a strong correlation of the items with the extracted component. This component accounts for 67% of the variance among the items. The other half of the data were then subjected to a confirmatory factor analysis, constraining the extraction to a single component, with similar results (loadings of .80 or greater and 72% of the variance explained).

With the eliminated pilot items included, the single derived component accounts for less of the variance among the items (63%) than the above models. Thus, the reduced size of the SRFI is warranted by this analysis.

Validity.

As stated earlier, the SRFI was developed using constructs having an empirically verified relationship with high quality instruction. This process gave us our best chance at having *construct validity*.

As a proxy for *concurrent validity,* we examined the relationship between our validation items and the SRFI items. The validation items were modeled after the constructs described in Feldman (2007) and concerns described throughout the SET literature. The phrasing of the validation items very closely matched corresponding items in the Instructional Development and

Effectiveness Assessment (IDEA; Hoyt & Cashin, 1977). We also tested against such variables as class size and average grade.

Relationship with Instructional Variables.
Some validation items represent variables that are at least somewhat under the instructor's influence. These include the extent to which students are challenged to do more than memorize material and the amount of knowledge students perceive they have gained within a course.

Quality of Instructional Goals.

A regression analysis with the quality of instructional goals (from the item "The instructor required students to do more than just memorize course material") as the dependent variable resulted in a model with 7 predictors and an R^2 of 0.568, $F(7, 755) = 141.752$, $p < .001$. All beta coefficients for the three SRFI items included in the model were positive, showing that SRFI ratings tend to be somewhat higher when the instructor requires students to do more than just memorize course material.

Knowledge Gain.

To analyze the relationship of SRFI items with student-perceived knowledge gain, a regression analysis was conducted with the item "I learned a lot in this course" as the dependent variable. This stepwise regression resulted in a model with 8 predictors and an R^2 of 0.611, $F(8, 753) = 148.096$, $p < .001$. All beta coefficients for the six SRFI items included in the model were positive, showing that SRFI ratings tend to be higher when students believe they have learned a lot in the course.

Relationship with Non-instructional Variables.

Some validation items represent variables that lie almost entirely outside of the instructor's control. If there is no relationship between responses to these factors and responses to SRFI items, it would be practically impossible to argue that these constructs are influencing SRFI ratings.

Student Motivation.
Our three validation items related to student motivation asked students about their (a) prior interest in the subject matter, (b) desire to enroll in a course with this instructor, and (c) experience in the course being pleasurable. For each of these items, students responded with the full range of possible values (0–4) across all courses in the aggregate; the aggregate median for these items, respectively, was (a) 3, (b) 2, and (c) 3. In a series of regression analyses with each of these validation items as the dependent variable, three SRFI items did not even enter the models. The other four SRFI items entered the models as predictors, sometimes at the first step, but never at the first step for all three models.
Not surprisingly, the conclusion is that we cannot rule out student motivation as a factor in SRFI ratings. However, we can also say that motivation does not appear to be the principal influence across the board.

Relationship with Other Variables.

There is some inconsistency in the SET literature as to whether class size and course grade make any difference in SET responses. To check this relationship with the SRFI, our third-party coordinator (to ensure anonymity) provided me with course-enrollment and student-performance data for each participating course.

Some articles in the SET literature report student-level correlations, and others report class-level correlations. Since there is no "correct" unit of analysis for SET investigations (Feldman, 2007; Wigington, Tollefson, & Rodriguez, 1989), both are provided (where appropriate) in Table 1. This is instructive, because the increased variability (and increased *n*) in the student-level correlations yields more statistically reliable results; in fact, only two of the 21 class-level correlations are statistically different from zero!

Class Size.

The typical finding in the SET literature—for those studies where any non-zero relationships are found—is that larger classes yield lower SET ratings (Feldman, 1984; Marsh, 1987). For the SRFI, this is somewhat true, although the correlations are quite weak to nonexistent. The only exception to this is SRFI item #6, which had a class-level correlation of -.49 with enrollment. As can be seen from the appendix, item #6 has to do with the instructor caring about the progress of each student in the course; therefore, a negative correlation is expected here on practical grounds alone.

Average Grade.

The relationship between student achievement and SET items is well established in the literature to be a correlation of about 0.30 or less (Feldman, 2007; Marsh & Roche, 1997). To look for relationships between SRFI data and student achievement (Feldman's first mechanism), the class average, as a percent, was correlated with each SRFI item. The resulting correlations with SRFI items are given in Table 1. As these correlations show, there is no significant relationship between SRFI ratings and the average grade for the course. Most importantly, there is no basis in these data for claiming that "easy" courses get higher ratings.

In this sample, the correlation between class size and average grade was -.46, $p < .01$. Thus, the larger classes tended to have lower average grades, which is not a surprising outcome given that the three largest classes in this sample were 100-level classes.

Course Level.

There is some indication in the SET literature that students in advanced courses provide ratings that are more favorable (Aleamoni, 1999; Cashin, 1995). The SRFI correlations related to this are in Table 1. There is at best a weak tendency for students in more advanced courses to rate their instructors more favorably. This is particularly good news for instructors of freshman classes, who often fear lower ratings from unengaged first-year students.

Pilot Study 2

Before recommending the SRFI for campus-wide adoption, it was subjected to a second pilot test. The purpose was to show that the results from Pilot Study 1 could be replicated.

Method

The procedure for Pilot Study 2 was identical to that of the first pilot study. We obtained results from 33 classes: 12 100-level, 12 200-level, 7 300-level, and 2 graduate classes. The median enrollment for these classes was 24 students.

Pilot Instrument 2

The instrument for the second pilot study was identical in form to Pilot Instrument 1. Pilot Instrument 2 retained the validation items from Pilot Instrument 1 but eliminated

the items that the regression analyses had not shown to be useful in Pilot Study 1.

Results and Discussion

Response Rate.

In this study, 638 students (87%) responded to the pilot instrument. This is a highly satisfactory response rate. From the 31 undergraduate classes, 616 students (86%) responded, and 22 students (96%) from the 2 graduate classes responded.

Regression Results.

Three stepwise regressions were again conducted, using the same dependent variables and criteria as in Pilot Study 1. Again, all regression models had omnibus *p*s less than 0.0001. The final models included between 5 and 6 predictor variables, all of which were SRFI items. The R^2 values for these final stepwise models ranged from 0.639 to 0.816; as in Pilot Study 1, we can account for a modestly strong proportion of the variance in the global validation items with the SRFI items.

Items Eliminated. There were no pilot items eliminated on the basis of the regression analyses. All of the pilot items retained from Pilot Study 1 functioned as expected.

Reliability.

Coefficient alpha (Cronbach, 1951) for the SRFI items was 0.92, practically identical to the finding from those same items in Pilot Study 1. The second reliability estimate calculated, ω_t, was 0.92, again identical to coefficient alpha and almost identical to the same metric in Pilot Study 1. The intraclass correlation for Pilot Study 2 was also 0.92. Thus, with virtually the same strong results across both of these pilot studies, the SRFI has shown itself to have impeccable reliability in these two samples.

Uni-Dimensionality.

A confirmatory factor analysis was conducted, constraining the extraction to a single component (due to the clear results of a single component from the factor analysis in Pilot Study 1). Once again only one component was extracted, with loadings of 0.748 and greater. These high loadings indicate a strong correlation of the items with the derived component. This component accounts for 68% of the variance among the items.

Validity.

The validity tests from Pilot Study 1 were repeated. These tests examined the relationship between SRFI responses and various course and student characteristics.

Quality of Instructional Goals.

The stepwise model for quality of instructional goals resulted in a model with 4 predictors and an R^2 of 0.349, $F(4, 633) = 84.997$, $p < .001$. All beta coefficients included in the model were again positive, showing that SRFI ratings tend to be higher when the instructor requires students to do more than just memorize course material.

Knowledge Gain.

The stepwise regression for student-perceived knowledge gain resulted in a model with 4 predictors and an R^2 of 0.520, $F(4, 633) = 171.322$, $p < .001$. All beta coefficients were again positive, showing that SRFI ratings tend to be higher when students believe they have learned a lot in the course.

Student Motivation.

The three stepwise regressions for student motivation were again conducted with the same three validation items as dependent variables: (a) prior interest in the subject matter; (b) desire to enroll in a course with this instructor; and (c) experience in the course being pleasurable. For each of these items, students responded with the full range of possible values (0–4) across all courses in the aggregate; the aggregate median for these items, respectively, was (a) 3, (b) 2, and (c) 3. The final stepwise models for these three regressions had the following characteristics, respectively: (a) 3 predictors and an R^2 of 0.08; (b) 3 predictors and an R^2 of 0.40; and (c) 4 predictors and an R^2 of 0.75.

Not surprisingly, we again see that student motivation is related to ratings. The only student-motivation item which the instructor has no control over is students' prior interest in the topic—and that accounts for only 8% of the variance in these data.

Class Size.

As shown in Table 2, the correlations with class size are very weak; most of them are not statistically distinguishable from zero. Therefore, class size does not appear to be influencing SRFI ratings within this sample.

Average Grade.

Students' average grade does not seem to be a heavy influence on SRFI ratings. Most of these relationships are weak, but two correlations bear further inspection. SRFI items #2 and #4 have a modest correlation with average course grade. However, neither of these items point to inflated ratings because of lenient grading. Item #2 is about the instructor demonstrating an interest in the course material; when this happens, it should come as no surprise that students in the class are achieving more. Item #4 has to do with the instructor effectively conveying why the subject is meaningful; would students not be expected to accomplish more when they have been told why the material is important?

In this sample, the correlation between class size and average grade was -.13, $p = .48$. The largest class size in this sample was 60 students, which could account for the nonsignificant finding compared to Pilot Study 1 in which the largest enrollment was 92 and there was a significant relationship between these variables.

Course Level. The correlations with course level shown in Table 2 are mostly weak or nonsignificant. The single exception is item #1, which has to do with the instructor explaining grading policies. The negative correlation indicates that instructors of higher-level courses did not accomplish this as well as their peers teaching lower-level courses. The higher-level courses in this sample likely incorporated more subjective forms of evaluation that were not transparent to students. However, there were only two graduate courses in this sample (in contrast to the five graduate courses in Pilot Study 1), so this correlation of -0.42 could be interpreted as an artifact of the sample when compared to the corresponding 0.08 correlation in Pilot Study 1.

General Discussion

Reliability

Three forms of reliability estimates were used: coefficient alpha, ω_t, and intraclass correlation. All three estimates were uniformly high across both pilot studies, demonstrating that—at least for these two samples—the SRFI is a highly dependable instrument.

One must be cautious, of course, in drawing conclusions beyond the sample data. Reliability is a characteristic of the data set,

not of the instrument itself. However, the similarity of all reliability coefficients across both pilot studies (along with the general findings of high reliability within the SET literature) does inspire confidence that reliability will not be a concern moving forward.

Validity

Validity remains the most critically important characteristic of any SET. Yet, as everyone acknowledges, establishing the validity of a SET is extraordinarily challenging. The difficulty stems from the fact that effective teaching is such a complex phenomenon that there exists no uniformly agreed upon, validated instrument for measuring it (Cashin, Downey, & Sixbury, 1994).

In the SET literature, the articles that describe the development of specific SETs typically depend upon expert opinion as the primary source of validation (e.g., Alok, 2011; Barnes et al., 2008; Kember & Leung, 2008). Our approach with the SRFI was to rely upon a broad empirical base of evidence rather than the opinions of a group of experts. This is not to be read as a slight against experts; neither is it intended to cast doubt upon their expertise. Rather, we understood that one of the chief difficulties of utilizing expert teachers for SET development is that they invariably disagree with one another. Additionally, it is not unthinkable that experts might change their minds over time. Therefore, using a body of empirical data that has been validated against outcomes important to effective teaching—like student performance on a final exam—over an extended period of time offers us an attractive alternative path to validity that is unique within the SET literature. One would certainly have to mount a considerable defense to convincingly demonstrate that all the studies Feldman (2007) analyzed were wrong! This approach to validation paid off. By establishing stringent criteria for inclusion of potential items, we gave ourselves the best chance at having a solid instrument on the first attempt. Indeed, the data from the SRFI items in both pilot tests match the pattern of results from Feldman (2007) perfectly. The evidence is unequivocal: Despite generational differences and technological advances, the students of today respond to instruction much like the students of yesterday did.

Interpretation

To help address McKeachie's (1997) concerns of invalid interpretation of SET data, my committee provided a list of guidelines for interpretation of SRFI data to the campus community. This list was especially intended for use by personnel committees when evaluating SRFI results for purposes of tenure and promotion.

Conclusion

After a careful, empirically driven process of instrument development, the SRFI performed admirably throughout two pilot tests. The reliability and validity (to the extent that validity could be measured) of this instrument in these two pilot studies were highly satisfactory. Already the SRFI has more evidence of reliability and validity than many instruments currently in use. Because the SRFI was piloted in many different disciplines, it can now be recommended for implementation on a broad scale. The relationship between SRFI ratings and course characteristics such as class size and level align with the expectations from the empirical literature. Consequently, the SRFI is a strong competitor to existing instruments—and should be preferred over instruments for

which validity has never been formally examined.

Future research on the SRFI should investigate how it functions on a campus-wide scale. Both faculty and administration on my own campus have approved it for campus-wide use as a replacement of our existing (non-validated) instrument; unfortunately, the faculty union leadership blocked its implementation. Therefore, no campus-wide data can be collected at my institution at present. Nonetheless, the SRFI remains a stronger instrument than many of its competitors.

References

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 385-456). Netherlands: Springer.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*, 153-166.

Alok, K. (2011). Student evaluation of teaching: An instrument and a development process. *International Journal of Teaching and Learning in Higher Education, 23*, 226-235.

Baldwin, T., & Blattner, N. (2003). Guarding against potential bias in student evaluations: What every faculty member needs to know. *College Teaching, 51*, 27-32.

Barnes, D. C., Engelland, B. T., Matherine, C. F., Martin, W. C., Orgeron, C. P., Ring, J. K., . . . Williams, Z. (2008). Developing a psychometrically sound measure of collegiate teaching proficiency. *College Student Journal, 42*, 199-213.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Cashin, W. E. (1995). *Student ratings of teaching: The research revisited* (IDEA Paper No. 32). Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology, 84*, 563-572.

Cashin, W. E., Downey, R. G., & Sixbury, G. R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: Reply to Marsh (1994). *Journal of Educational Psychology, 86*, 649-657.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education, 21*, 45-116.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Netherlands: Springer.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*, 121-135.

Hoyt, D. P., & Cashin, W. E. (1977). *Development of the IDEA system* (IDEA Technical Report No. 1). Manhattan, KS: Center for Faculty Evaluation and Development, Kansas State University.

Kember, D., & Leung, D. Y. P. (2008). Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education, 33*, 341-353.

Kidwell, J. S., & Brown, L. H. (1982). Ridge regression as a technique for analyzing models with multicollinearity. *Journal of Marriage and Family, 44*, 287-299.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43*, 711-742.

Lindahl, M. W., & Unger, M. L. (2010). Cruelty in student teaching evaluations. *College Teaching, 58*, 71-76.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523-553.

Marsh, H. W. (1982a). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*, 77-95.

Marsh, H. W. (1982b). The use of path analysis to estimate teacher and course effects in student ratings of instructional effectiveness. *Applied Psychological Measurement, 6*(1), 47-59.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Netherlands: Springer.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*, 1187-1197.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 3-23.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*, 1218-1225.

McKeachie, W. J. (2007). Good teaching makes a difference—and we know what it is. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 457-474). Netherlands: Springer.

McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1173-1190.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ratcliff, R., & McKoon, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2*, 754-767.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*, 145-154.

Schacter, D. L., & Cooper, L. A. (1995). Bias in the priming of object decisions: Logic, assumption, and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 768-776.

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120.

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales. *Assessment & Evaluation in Higher Education, 32*, 667-679.

Stapleton, R. J., & Murkison, G. (2001). Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and expected grades. *Journal of Management Education, 25*, 269-291.

Wigington, H., Tollefson, N., & Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education, 30*, 331-344.

Table 1

Pearson Correlations Between SRFI Items and Class Characteristics in Pilot Study 1

Item[a]	Class Size[b]		Avg. Grade[c]	Course Level[d]
	Student[e]	Class[f]	Class	Class
Item 1	0.00	-0.02	-0.12	0.08
Item 2	-0.29**	-0.27	-0.04	0.08
Item 3	-0.03	-0.08	0.04	0.07
Item 4	-0.27**	-0.33*	0.11	0.19
Item 5	-0.16**	-0.21	0.14	0.16
Item 6	-0.36**	-0.49**	0.11	0.30
Item 7	-0.14**	-0.23	0.17	0.19

Note. No student-level data are available for course grade or class year; therefore, only class-level correlations are shown for these variables.

[a]See the appendix for full item text. [b]Range: 2 to 92 students enrolled [c]Range: 71% to 96%

[d]Course Level was coded as follows: 1 = 100-level course; 2 = 200-level course; 3 = 300-level course; 4 = graduate course (our institution does not have 400-level courses). [e]Individual-level correlations between SRFI items and class size. [f]Class-level correlations between SRFI items and class size.

*$p < .05$. **$p < .01$.

Table 2

Pearson Correlations Between SRFI Items and Class Characteristics in Pilot Study 2

Item[a]	Class Size[b]		Avg. Grade[c]	Course Level[d]
	Student[e]	Class[f]	Class	Class
Item 1	0.12**	0.31	-0.04	-0.42*
Item 2	0.02	-0.03	0.51**	0.26
Item 3	0.04	0.18	0.23	-0.17
Item 4	-0.07	-0.16	0.47*	0.31
Item 5	0.03	0.11	0.15	-0.18
Item 6	-0.03	0.05	0.36	0.08
Item 7	0.09*	0.19	0.18	-0.26

Note. No student-level data are available for course grade or class year; therefore, only class-level correlations are shown for these variables.

[a]See the appendix for full item text. [b]Range: 2 to 60 students enrolled [c]Range: 75% to 94%

[d]Course Level was coded as follows: 1 = 100-level course; 2 = 200-level course; 3 = 300-level course; 4 = graduate course (our institution does not have 400-level courses). [e]Individual-level correlations between SRFI items and class size. [f]Class-level correlations between SRFI items and class size.

*$p < .05$. **$p < .01$.

Appendix

Student Response to Faculty Instruction

1. The instructor clearly explained his/her grading criteria, including how final grades in this course will be determined.
2. The instructor was clearly interested in the course material.
3. The instructor presented and explained ideas effectively.
4. The instructor communicated the significance of the subject.
5. Throughout the course, the instructor made it clear what I should learn and accomplish.
6. The instructor was clearly interested in the learning of each student.
7. I would recommend this instructor to other students.

Team Performance Pay and Motivation Theory: A Mixed Methods Study

Author Info:

Pamela Wells, EdD, Executive Director, Region 4 Education Service Center, 7145 West Tidwell Road, Houston, Texas 77092, Email: pam.wells@esc4.net, 281-832-7786

Julie P. Combs, EdD Sam Houston State University, Educational Leadership and Counseling Department, Box 2119, Huntsville, TX 77341, Email: jcombs@shsu.edu, Phone: 936 294 3181

Rebecca M. Bustamante, PhD, Sam Houston State University, Department of Educational Leadership, Box 2119, Huntsville, TX 77341, bustamante@shsu.edu (932) 294-4946

Abstract

This study was conducted to explore teachers' perceptions of a team performance pay program in a large suburban school district through the lens of motivation theories. Mixed data analysis was used to analyze teacher responses from two archival questionnaires (Year 1, n = 368; Year 2, n = 649). Responses from teachers who participated in the team pay performance system reflected high levels of expectancy. Results were mixed for teachers' perceptions of equity. Some teachers expressed concerns related to distributive justice and procedural justice of the performance pay process. Implications for researchers and practicing educators are discussed.

Keywords: Teacher performance, Team performance pay, Performance pay, Motivation theory

Team Performance Pay and Motivation Theory: A Mixed Methods Study

In recent years, an increasing number of U.S. school districts have implemented teacher performance pay programs to retain teachers and motivate them to improve student achievement. Politicians, business leaders, and educational reformers have promoted teacher performance pay, as evidenced by a number of programs being implemented across the country (Delisio, 2012; Podgursky & Springer, 2007). However, a dearth of empirical studies exists to support the increase of these programs. As such, more research about performance pay is needed (Lavy, 2002; Podgursky & Springer, 2007). The purpose of this study was to explore teachers' perceptions of a team performance pay program in a large suburban school district through the lens of motivation theories (i.e., expectancy, [Vroom, 1964]; and equity [Adams, 1965]).

Review of the Empirical Literature

Taylor and Springer (2009) defined performance pay programs as measures to reward teachers "for the additional effort" invested in effective teaching (p. 3). Essentially, performance pay programs are designed to motivate teachers to improve students' academic achievement. Because performance pay has been detrimental to collaboration among teachers, team performance pay systems have been implemented in recent years (see Springer, 2010; Springer & Balch, 2010).

Performance Pay Issues in Schools

Performance or merit pay in schools has been fraught with controversy. Some educator groups have argued against the merits of performance pay for teachers (Baker et al., 2010). The two largest teacher organizations, the National Education Association (NEA) and the American Federation of Teachers (AFT), opposed the inclusion of performance pay experiments in the No Child Left Behind Act (NEA, 2011). They claimed that the focus on performance pay obscured the real problem of inadequate teacher salaries. More recently, the unions have moderated their strong stance against performance pay, and NEA (2011) has expressed its support if districts utilizing such programs agreed to use collective bargaining processes. Some researchers have attributed the lack of success of performance pay systems to an unfair reliance on standardized testing as the basis for pay (Eberts, Hollenbeck, & Stone, 2002) and to internal dissension among educators caused by perceived inequities in award distribution (Murnane & Cohen, 1986).

Team Performance Pay and Teacher Collaboration

To counter the criticisms of individual performance pay systems, team performance pay approaches, called school-based performance awards (SBPA), have been implemented to promote teacher collaboration (Odden & Kelley, 2002). Researchers focused on motivation implications of SBPAs have suggested important design considerations (Kelley, Heneman, & Milanowski, 2002). Teachers must believe they will actually receive the performance award if earned. This expectation is problematic when school systems might not have the resources to insure sustainability of SBPAs. Moreover, teacher expectancy is critical. Teachers must believe the programs are fair and that goals are attainable. Kelley et al. (2002) concluded that "motivational impact is not guaranteed simply by promising teachers a bonus" (p. 397).

Several problems associated with team performance pay are as follows: (a) measurement problems (e.g., fair and accurate evaluations); (b) negative effects on collegiality; (c) unintended consequences (e.g., cheating); (d) increased costs; (e) union opposition; and (f) past failures of performance pay systems (Lavy, 2007). Additionally, *free riders*, or teachers who receive awards based on the efforts of their team members, were identified as a problem (Eberts et al., 2002; Lavy, 2007). Strategies to overcome these obstacles include structuring group incentives, encouraging peer pressure, and monitoring free riders (Lavy, 2007). In conclusion, team awards have encouraged teacher collaboration without many of the inherent problems of perceived fairness in individual performance pay programs (Raham, 2000).

Conceptual Framework

The conceptual framework for this study was based on expectancy theory and equity theory, which informed the study in a variety of ways. First, these theories were used to organize the data during the data analysis phase. Second, the theories were utilized to understand the motivational benefits of a performance pay system in schools, particularly in making sense of teachers' perceptions during the data interpretation phase. Finally, when applying the concepts of expectancy and equity to school settings, several key motivation concepts were considered in drawing conclusions from the results.

Expectancy

An important and applicable motivational construct is expectancy theory. Vroom's (1964) expectancy theory suggests that employees will intentionally choose to put forth effort if they believe their efforts will be realized and rewarded with something that they perceive will meet their personal goals. Vroom's theory describes a motivational process rather than a cause and effect relationship (Fudge & Schlacter, 1999). Quick (1988) outlined the following steps for enacting expectancy theory: (a) define expectations, (b) make work valuable and doable, (c) give regular feedback, and (d) reward employees when they meet expectations. Isaac, Zerbe, and Pitt (2001) summarized Vroom's description of expectancy theory as a way people determine behavior based on their individual perceptions in order to increase pleasure and reduce pain. When evaluating the relationship between expectancy theory and incentive programs, Mathibe's (2008) research indicated that respondents' low levels of motivation were due to the absence of incentive programs to reward performance.

Equity Theory

Equity theory relates to how individuals continually assess the personal return they receive for the investment they put into the organization. Adams (1965) originally postulated that the extent to which people believe they are being treated in an equitable and fair manner could profoundly affect their motivation. Adams (1965) suggested that a motivational tension was created when a worker senses inequity when comparing one's inputs (effort) and outcomes (rewards) with those of other workers. Workers perceiving inequities might have feelings of anger or guilt. Further, a perceived lack of equity often serves as a de-motivator for workers (Mathibe, 2008). Robbins and Judge (2009) posited that when employees perceive inequity, they could be predicted to react in a variety of ways including exerting less effort, changing their rates of productivity, altering their perceptions of

others, or even quitting their jobs. Likewise, Milanowski (2000) proposed that teachers' perceptions of fairness were related to the "motivational power of monetary performance rewards" (para. 19).

In summary, these motivational theories provided a framework in this study for evaluating teachers' perceptions about a team performance pay system. Key concepts in these theories included the extent to which teachers believed their efforts would be rewarded and the perceived fairness of the performance pay process.

Method

Teachers' perceptions of the team performance pay program were elicited using mixed-item questionnaires that were administered over a period of two consecutive academic years. Mixed methods techniques were used in the data integration and data interpretation stages (Onwuegbuzie & Teddlie, 2003). Results were interpreted through the lens of the motivational theories described in the previous section.

Sample

The sample included all teachers from at-risk elementary schools and middle schools that implemented the District Awards for Teacher Excellence (DATE) team performance pay program in the selected school district. The DATE program schools were Title I campuses with high percentages of students in poverty. In the first year, 368 teachers represented 13 elementary schools and six middle schools in the program. In the second year of implementation of the team performance pay program, 649 teachers participated from 14 elementary schools and seven middle schools. More teachers were included in Year 2 because of student growth in the schools and the addition of two new schools.

All schools were located in a large, fast growing urban/suburban school district with a student enrollment of 106,000 students. The DATE program lasted for three years and was discontinued due to state funding shortages. For this school district, DATE was designed to increase teacher retention on at-risk campuses and to motivate teachers to collaborate for improved student performance.

Instrumentation

Archived data were obtained from the National Center on Performance Incentives (NCPI) at Vanderbilt University Peabody College. Data were collected by NCPI researchers using questionnaires to measure attitudes about DATE, the school environment, and professional practices (Springer et al., 2010). Some of the questions were based on prior, validated surveys including the Schools and Staffing Survey and the Consortium on Chicago School Research. The questionnaire included 33 questions divided into the following sections: (a) professional title, (b) attitudes about the DATE program, (c) school environment, (d) curriculum and instruction practices, (e) background information and teacher compensation information. The survey had been administered to teachers in the NCPI Nashville study and in the evaluation of the Texas' performance pay programs; therefore, it had been subjected to tests for internal reliability (J. Lewis, personal communication, April 28, 2011). For this study, we analyzed 55 common items (closed and open-ended responses) from the NCPI questionnaires.

Data Analysis

Qualitative, quantitative, and mixed analysis techniques were used. First, using an *a priori* approach, all questionnaire items were categorized into thematic clusters that most reflected the motivational theories applied in the study. Using a constant comparative approach established by Strauss and Corbin (1990), this sorting was conducted by reading each item and coding items most related to the major tenets of each theory. To increase the qualitative research validity (Johnson, 1997), an expert peer reviewer was asked to code the items separately, and the initial intercoder agreement rate was 94%. Next, descriptive statistics were generated for each of the closed-item responses.

Conclusions were drawn using a mixed data analysis approach (Combs & Onwuegbuzie, 2010). Quantitative and qualitative data were mixed using cross-over analysis strategies, specifically that of data integration (Onwuegbuzie & Teddlie, 2003). A parallel mixed analysis process was used, which involves the independent analysis of quantitative and qualitative data. Finally, the data were combined in the data interpretation stage where inferences were constructed.

Findings

Mixed methods analysis of quantitative and qualitative data from the questionnaires revealed several results related to teachers' perceptions of the team performance pay program. These results were examined through the lens of the conceptual framework and are organized using these motivational theories.

Expectancy

In both years of the study, teachers expressed high levels of expectancy related to performance pay. They believed that their efforts would lead to an award, that they would be rewarded for their performance, and that they would meet their personal goals. For example, most teachers (78.4%) in Year 1 responded that they had a clear understanding of the criteria to earn an award. Even more teachers in Year 2 (90.6%) claimed a clear understanding. Teachers believed that criteria established in the incentive plan were worthy of extra pay, with an 85.2% agreement rate. In Year 1, 56.5% of the teachers expected to receive an award, increasing to 78.8% in Year 2.

Additional insights related to expectancy theory were gained from open-ended comments. One teacher described the motivational benefit of a goal focus to earn an award: "The pay incentive has been an excellent tool in motivating me to take more staff development classes." However, another teacher provided a clearly articulated statement of negative expectancy: "From what I have seen so far is that the goals set by the program are not reachable nor is it enough money to motivate a teacher to go beyond what is already required of them."

Equity

Research findings were mixed for the category of equity (i.e., distributive justice, organizational justice, and procedural justice). The quantitative data suggested high levels of equity, but the qualitative data reflected concerns, especially about distributive justice (i.e., fairness of award distribution).

Teachers believed that the incentive plans used in their schools were fair (i.e., 80.3% in Year 1; 71.1% in Year 2). The percentage of agreement in Year 2, although still substantially high, was 9.2% lower than Year 1 results. In both years, the qualitative data reflected concerns about equity, particularly related to the rules for allocating awards. Statements related to a lack of fairness were more prevalent in Year 2 even though more teams (81% vs. 64%) received an award the second year. In addition, most teachers believed that the award amounts were large enough to be motivating. A reverse-coded item identified that a small number of teachers agreed that the award was not large enough to be motivating to them (i.e., 17.5% in Year 1; 16.4% in Year 2).

The majority of the open-ended comments expressed teachers' discontent about the perceived uneven distribution of awards. The number of comments in this category increased during Year 2, perhaps because teachers became aware of the differences in award amounts. Comments frequently reflected perceptions of a lack of distributive and procedural justice. According to one teacher, "The difference in the amount of money available to teachers is so great that those of us who do not teach an academic class do not feel that our position really holds any value."

Discussion

The majority of teachers surveyed in both years reported positive overall perceptions of team performance pay. The results provided evidence of teachers' motivational beliefs related to the pay program and their efforts to improve student outcomes. Some data suggested an increase in teacher agreement rates in Year 2, with more teachers expecting awards and believing that their efforts mattered. .

Much of the evidence focused in the area of equity theory (Adams, 1965; Mathibe, 2008), which related to teachers' perceptions of distributive justice (i.e., fairness of award amount and distribution), organizational justice (i.e., fairness of the workplace), and procedural justice (i.e., fairness of the process). In responding to close-ended items, teachers expressed beliefs that the program was equitable. However, in the open-ended responses, many teachers commented on their perceived lack of equity about unfair award distribution. Teachers' concern about distributive justice in performance pay programs has been reflected in the literature as well (e.g., Mahony, Menter, & Hextall, 2004; Murnane & Cohen, 1986).

In regards to distributive justice, teachers at at-risk campuses who taught in grade levels that were tested as part of the state accountability system were eligible to receive substantially higher awards than other teachers. Yet, the program was designed to reward teacher teams as a means to enhance equity. Proponents have reasoned that team award systems address the concerns of competition among teachers and the uneven distribution of difficult students. Results from this study indicated that this team component enhanced perceptions of equity for teachers who taught state-tested content courses. However, perceptions of equity were diminished for those teachers who did not have the opportunity to receive as large an award because they taught subjects that were not state-tested.

Conclusion

One of the strengths of this particular study about teacher team performance pay was that quantitative and qualitative data were analyzed. As Tashakkori and Teddlie (2009) point out, mixed methods can be more useful in determining if evaluation criteria are met, because stronger inferences can be made from the breadth of information provided in numbers and the depth of information given in narrative responses. In this research, the sampled teachers were able to express their overall perceptions of the program in response to the closed-response items on the surveys as well as describe their feelings and opinions about specific aspects of the program. Therefore, the analysis of both the quantitative and qualitative data provided a clearer picture of teachers' perceptions of the performance pay program in a way that can inform future research, implementation of teacher performance pay programs, and district and state policies related to performance pay in education.

Although we were able to examine data related to a team performance pay system with a large sample of teachers, limitations remain. A possible limitation was that the findings from the study might not be generalized to other settings. Therefore, to minimize that possibility, we utilized strategies to enhance the rigor of the research including triangulation of qualitative and quantitative research. Another limitation was that some schools did not respond in one of the two survey years. In addition, because the data were archival, the instrument was already designed, and this limitation did not allow any control over the variables or themes included in the survey. As a result, the survey did not take into account other aspects of team performance pay systems.

Several recommendations for implementing teacher performance pay systems became apparent. Teachers' understood the program and were motivated to improve student performance; therefore, administrators might need to provide clearer and more consistent information about program structures and the necessary criteria to achieve awards. As one teacher expressed, "teachers are not able to meet goals they don't know about." We suggest that practitioners plan for multiple methods of communication so that teachers understand the program rules and remain motivated to achieve the awards.

Another practical recommendation relates to the notion of teacher team motivation. Although teachers in this study reported enhanced cooperation within their respective teams, some teachers who taught in subjects or grades that were not tested believed that the performance system was unfair, inequitable, and not motivating. Consequently, educational leaders should be aware that even though team performance pay might reinforce cooperation within teacher teams, the criteria also could cause division among teachers in the school. District and campus leaders should consider ways to measure the contributions of teachers who do not teach subjects tested in state assessments. Additionally, policymakers might consider whether performance pay programs should allow teachers at all grade levels in all academic subject areas to have the opportunity to earn monetary awards of equal values.

Finally, further study about teacher performance pay is needed, particularly as more U.S. school districts implement such programs. One question that remains is if money is the motivator for teachers to improve their instructional strategies and to collaborate to improve student achievement. Policymakers and district leaders should consider how these limited funds are best used and distributed. Furthermore, more

research is needed on the potentially negative impact that performance pay could have on the motivation of teachers who teach subject areas that are not eligible for performance awards.

References

Adams, J. S. (1965). Inequity in social exchanges. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 267-300). New York, NY: Academic Press.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . & Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. *EPI Briefing Paper 278.* Retrieved from http://www.epi.org/publication/bp278/

Combs, J. P., & Onwuegbuzie, A. J. (2010). Describing and illustrating data analysis in mixed research. *International Journal of Education, 2,* 1-23. Retrieved from http://www.macrothink.org/journal/index.php/ije/article/view/526

Delisio, E. R. (2012). Pay for performance: What are the issues? *Education World*. Retrieved from http://www.educationworld.com/a_issues/issues/issues374a.shtml

Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcome. *The Journal of Human Resources, 37,* 913-927. doi:10.2307/3069621

Fudge, R. S., & Schlacter, J. L. (1999). Motivating employees to act ethically: An expectancy theory approach. *Journal of Business Ethics, 18,* 295-304. doi:10.1023/A:1005801022353

Isaac, R., Zerbe, W., & Pitt, D. (2001). Leadership and motivation: The effective application of expectancy theory. *Journal of Managerial Issues, 13,* 212-226.

Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education, 118*, 282-292.

Kelley, C., Heneman, H., & Milanowski, A. (2002). Teacher motivation and school-based performance awards. *Educational Administration Quarterly, 38,* 372-401. doi:10.1177/0013161X02383004

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *The Journal of Political Economy*, *110,* 1286-1317. doi:10.1086/342810

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The Future of Children*, *17,* 87-109. doi:10.1353/foc.2007.0007

Mahony, P., Menter, I., & Hextall, I. (2004). The emotional impact of performance-related pay on teachers in England. *British Educational Research Journal*, *30*, 435-456. doi:10.1080/0141192041000168972 4

Mathibe, I. (2008). Expectancy theory and its implications for employee motivation. *Academic Leadership*, *6*, 2. Retrieved from http://www.academicleadership.org /emprical_research/466.shtml

Milanowski, A. (2000). School-based performance award programs and teacher motivation. *Journal of Education Finance, 25*, 517-544.

Murnane, R., & Cohen, D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Education Review, 56,* 1-17.

National Education Association. (2011). *Fact sheet on NEA's views regarding mandated performance pay for educators*. Retrieved from http://www.nea.org/home/15069.htm

Odden, A., & Kelley, C. (2002). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools* (2nd ed). Thousand Oaks, CA: Corwin Press.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.

Podgursky, M., & Springer, M. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, *26*, 909-949. doi:10.1002/pam.20292

Quick, T. (1988). Expectancy theory in five simple steps. *Training and Development Journal, 42*, 30-32.

Raham, H. (2000). Cooperative performance incentive plans. *Peabody Journal of Education, 75*, 142-158. doi:10.1207/S15327930PJE7504_7

Robbins, S., & Judge, T. (2009). *Organizational behavior.* Upper Saddle River, NJ: Pearson Prentice Hall.

Springer, M. G. (Ed.). (2010). *Performance incentives: Their growing impact on American K-12 education*. Washington, DC: Brookings Institution.

Springer, M.G., & Balch, R. (2010). Design components of incentive pay programs in the education sector. In S. Sclafani (Ed.), *Teacher Incentives and Stimuli* (pp. 65-94). Paris: Organisation for Economic Co-Operation and Development. Retrieved from https://my.vanderbilt.edu/performanceincentives/files/2012/10/Springer_Balch_OECD_Chapter1.pdf

Springer, M. G., Lewis, J. L., Ehlert, M. W., Podgursky, M. J., Crader, G. D., Taylor, L. L., ... Stuit, D. A. (2010, December). *District Awards for Teacher Excellence (D.A.T.E.) program: Final evaluation report.* Austin, TX: Texas Education Agency. Retrieved from Texas Education Agency website: http://www.tea.state.tx.us/index4.aspx?id=2928&menu_id=949

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Newbury Park, CA: Sage.

Tashakkori, A. & Teddlie, C. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Thousand Oaks, CA: Sage.

Taylor, L. L., & Springer, M. G. (2009). *Optimal incentives for public sector workers: The case of teacher-designed incentive pay in Texas.* Nashville, TN: National Center on Performance Incentives.

Vroom, V. (1964). *Work and motivation.* New York, NY: Wiley.

Why aren't they paying attention to me? Strategies for preventing distraction in a 1:1 learning environment

Author Info:

Jennifer T. Tagsold, Ph.D., jtagsold@peace.edu

Abstract

K-12 education research has become increasingly concerned with technology's impact on students' attention in the classroom, particularly with regard to laptop computers and other mobile devices (Gay & Hembrooke, 2004; Jackson, 2008; Mann, 2008; Kraushaar & Novak, 2010). While this classroom technology has created many positive implications for teaching and learning, few scholars have examined specific, practical approaches teachers use to avoid distraction. Through qualitative data analysis methods, this study examines the strategies used by teachers to avoid distraction in a high-tech learning environment.

INTRODUCTION

Sam had been using his new laptop for two weeks. He appreciated the shiny new tools available to him and got a kick out of watching his teachers figure out how to use technology, but he found himself constantly distracted. "Here Mr. Williams, I'm returning my laptop because I'm not getting any work done; if it's not here to distract me, I will be able to pay more attention to what you say in class."

Mr. Williams enjoyed having total access to myriad teaching tools through the new 1:1 laptop initiative, but he also had to learn to monitor his students. As his teaching became more tech-savvy and more directed at 21st century skill development, Mr. Williams developed a new goal: become more interesting than all the distractions available on his students' laptops. After all, if Facebook was more compelling than his lesson, what kind of teacher was he? Surely he could compete with social networking and other online activities…but how?

As a research assistant on a project that evaluates the 1:1 computing initiative in North Carolina schools, I learned that most students easily adapt to the use of laptops in the classroom and experience positive effects on their entire learning experience due to laptop access (Bebell, 2005; Corn, 2009; Great Maine Schools Project, 2004; Lane, 2003; Lowther et al., 2007; Silvernail & Lane, 2004; Warschauer, 2006). One primary issue that kept arising, however, was that of distraction. This paper seeks to enhance the field of instructional technology by exploring strategies that teachers use to maintain a focus on teaching and learning in a 1:1 environment when so many other options are available to distract the minds of the learners.

LITERATURE REVIEW

Benefits and Drawbacks of 1:1 Learning

Because of digital media's impact on learning styles, researchers must examine the fruits of laptop initiatives and how these help student navigate the non-linear world of the Internet and its similarities to the "associational network of human long-term memory" (Dede, 2005, p. 5). A discussion of the positive and negative components of 1:1 learning will help reveal some of these fruits. Positive outcomes of 1:1 learning include enhanced student learning and engagement (Bebell, 2005; Mitchell Institute, 2004; Lane, 2003; Lowther, Strahl, Inan, & Bates, 2007; Silvernail & Lane, 2004; Warschauer, 2006), motivation (Harris & Smith, 2004; Mitchell Institute, 2004; Silvernail & Lane, 2004), achievement (Cavanaugh, Dawson, White, Valdes, Ritzhaupt, & Payne, 2007; Lowther et al., 2007), attendance (Harris & Smith, 2004; Lane, 2003; Silvernail & Lane, 2004), discipline (Corn, 2009; Silvernail & Lane, 2004;), and 21st century skills (Cavanaugh et al., 2007; Corn, 2009; Lowther et al., 2007; Mitchell Institute, 2004; Shapley et al., 2008). Results from 1:1 initiatives have also shown an increase in students' math and writing skills (Bebell, 2005; Warschauer, 2006). Classroom teachers report benefits of 1:1 learning such as improved technology knowledge and skills, increased assistance with technology questions and problems, and improved classroom management (Fairman, 2004).

Although some results of 1:1 laptop initiatives show improvement in student learning and teacher use, factors other than the distribution of laptops contribute to successful implementation. There is

evidence that laptop initiatives do not increase all test scores, especially when tests are administered in paper and pencil form (Warschauer, 2006; Weston & Bain, 2010). Other drawbacks of 1:1 learning include distraction and reduced attention (Mann, 2008), student misuse (Holcomb, 2009), physical discomfort leading to a need for ergonomics training (Fraser, 2002), and lack of teacher and student technical skills (Corn, Halstead, Tingen, Townsend, & Campbell, 2010). Teacher support, instructional use, technology support, infrastructure, and quality of implementation are influential in the success of a 1:1 initiative (Weston & Bain, 2010). In 1:1 laptop initiatives, students are provided laptops for educational use; however, the schools must have the capabilities and strategies for the laptop use to be effective (Warschauer, 2006). This includes technology support, resources, and strong leadership guiding the programs (Kleiger Ben-Hur & Bar-Yossef, 2010; Maninger & Holden, 2009; Silvernail & Lane, 2004).

Improved Teaching, Depending on …

Teachers' beliefs mediate the way they use technology in the classroom, and if teachers do not support the initiative they are less likely to integrate the laptops into their lesson plans (Antonietti & Giorgetti, 2006; Churchill, 2006; Ertmer, Addison, Lane, Ross &Woods, 2000; Penuel, 2006). In addition to school and district support, teachers should support laptop learning in the classroom and have access to professional development or tools to aid them in integrating laptops into lesson plans (Kleiger Ben-Hur & Bar-Yossef, 2010; Penuel, 2006; Silvernail & Lane, 2004; Weston & Bain, 2010).

Self-Regulation Theory and 1:1 Learning

Self-regulation is defined as "how a person exerts control over his or her own responses so as to pursue goals and live up to standards" (Baumeister and Vohs, 2004, p. 500). A famous study involving children and marshmallows (Mischel, Ebbesen, & Raskoff Zeiss, 1972) helps to shed light on an important component of self-regulation: willpower. Mischel left children alone in a room with a bell and a promise that if they rang that bell before 20 minutes were up, they would receive one marshmallow; if they waited until 20 minutes had gone by, they could have two marshmallows. Mischel et al. found that participants were able to wait when they distracted themselves from the rewards (1972). Metcalfe & Mischel (1999) suggest that willpower consists of a cool, cognitive "know" system and a hot, emotional "go" system which affects the way humans perceive the world; stress, developmental level, and one's self-regulatory dynamics determine the balancing act between hot and cold labels. Students may be experiencing so much distraction in 1:1 initiatives due to tempering certain online activities as hot (i.e. Facebook) and others as cold (in-class assignment). Vohs & Baumeister (2004) assert that managing attention may be the most effective approach to self-regulation. Posner's theory of attention was used to guide the author's understanding of humans' ability to pay attention in learning environments. Posner asserts that the orienting network is the flashlight that directs our focus; the alerting network relates to wakefulness; and the executive network is at the heart of controlling attention (and oneself) (Posner & Boies, 1971; Posner & Rothbart, 2007).

How Students Avoid Distraction in High-Tech Learning Environments

In *Brain Rules*, John Medina suggests that a key brain rule is that humans do not pay attention to boring things (2008). This rule is evident when observing a student being torn between his or her laptop and focusing on the teacher lecturing at the front of the room. In one study, students with laptops spent considerable time multitasking and the laptop posed a significant distraction to users and fellow students; laptop use was negatively related to academic success (Fried, 2008). Another study found that students forced to close their laptops during a lecture were able to recall more lecture content than those who were not. Interestingly, the length of browsing time is extremely important; a lengthy browsing time "appears to be the nemesis of the multitasker; if one is adroit at staccato-like browsing, processing multiple inputs simultaneously may not suffer to the same extent" (Hembrooke & Gay, 2003, p. 59).

Various methods have been employed to help students avoid distraction in high-tech learning environments (Johnson, 2010). Some instructors found that making use of other technologies, like classroom response systems (clickers) and music, reduced distraction (Cole, 2010; Johnson, 2010). Some educators emphasize setting ground rules for wireless use at the beginning of the semester, which include students remembering their role as learners in class, not continually checking e-mail or instant messaging during class, and not handling the "business side" of life during class (*Wireless in the Classroom: Advice for Students*, 2011). Johnson (2010) suggests dealing with distraction by developing rules for laptop usage in collaboration with students; using technology to enhance traditional teaching (e.g., have students create a video instead of writing an expository paper); walking around the classroom periodically to monitor students; and using the technology to restructure the educational process.

Monitoring Software

Numerous schools involved in 1:1 initiatives use monitoring software to ensure that students stay focused on the lesson. Popular types of monitoring software include Eduplatform, E-Chalk, and DyKnow, which offer features such as group chat, teacher viewing of each student's computer, teacher viewing of students while taking tests, and freezing every computer to gain students' attention (Donnalley, 2011). Teachers have commented on DyKnow's ability to hold students accountable, engage students every day, increase teaching time, and energize the classroom (Donnalley, 2011). Some issues occur when monitoring software blocks sites that are useful for teaching and learning.

Instructional Practice

English teachers use their laptops for innovative instructional practices, such as digital storytelling (Lambert, 2002); multimodal texts to enhance reading comprehension (McKenna, 1998); creating student election commercials (Curtis, Merry, & Walker, 2011); and Google Docs for improved writing (Pahomov, 2011). Having digital literacy involves numerous skills, including knowing how to explore the Internet, find necessary information, and share that information with others (Leu, Leu, & Coiro, 2004).

RESEARCH QUESTION AND METHODOLOGY

Research Question

The primary research question under investigation is: How do English IV teachers help students manage distraction when working with technology?

Research Design

Because the researcher seeks to discover ways in which teachers and students manage distraction, a qualitative orientation to the project was preferable over quantitative research and its reliance upon predetermined evaluation instruments. This study follows the multiple-case study design, a commonly used method for "a study of school innovations…in which individual schools [or classrooms within a school] adopt some innovation" (Yin, 2003, p. 46). The primary research question was answered through qualitative methods, including interviews and classroom observations completed on location in the school environment.

Participants

High school students and teachers from one western North Carolina school involved in a 1:1 initiative participated in the study. This school is a recipient of the IMPACT grant, which provides schools with a model for technology integration, with components such as: having a full-time technology facilitator and media coordinator in place; developing a school-wide focus on flexible access to computer labs, mobile computer carts, and libraries; a 1:1 setting in some cases; and collaborative planning (Mollette et al., 2011). Because the chosen school does not use school-based monitoring software, the researcher was able to get a true sense of participants' strategies for managing distraction. At least 20 total interviews (sixteen students, four teachers) and eight classroom observations in both honors and traditional English language arts classes were conducted for this study.

Data Sources and Instruments

The study was conducted in three phases. Phase I and III consisted of classroom observations, using an updated version of the LoFTI (Looking for Technology Integration) observation instrument, and Phase II consisted of interviews. Each teacher was observed teaching twice, and each teacher and student participant was interviewed once. After the final classroom observations were completed, a debriefing occurred with each teacher to ensure member checking of the researcher's interpretations (Miles & Huberman, 1994).

Data Analysis Procedures

Interviews were analyzed using thematic content analysis (Corbin & Strauss, 2008) to answer the research question along with the sub questions. The researcher began with a short list of tentative codes that match text segments (Creswell, 2007) and then organized strategy codes, comprised of methods used to accomplish a goal, into categories (Bogdan & Biklen, 2007). The researcher then utilized the statements that best illustrate these categories to understand how each participant avoids (or helps others to avoid) distraction in a 1:1 learning environment (Hesse-Biber, 2010). The researcher developed codes until the point of saturation was reached (Denzin & Lincoln, 2000).

RESEARCH FINDINGS

How Teachers Help Students to Avoid Distractions

When asked how they help students deal with online distractions at school, teachers reveal several unique solutions. The primary way teachers help their students is to create lessons and assignments that are so engaging that students do not desire any outside

stimulation. Along with keeping the lesson engaging, teachers seek to keep their students so busy that they do not have time to seek distractions online. One key way teachers do this is through deadlines; when a student sees the deadline looming, particularly in discussion boards when his or her peers have already posted their assignments, he or she focuses on the task at hand. Teachers find that walking around the room and monitoring what students are doing helps them to stay on task. One teacher mentions her goal of teaching from "bell to bell," using a system in which students move on to another task if they have completed the first one; when finished, they are permitted to complete homework from other classes.

When asked about the processes students use to complete assignments on their laptops, teachers reveal a variety of methods. Students typically submit assignments to an online drop box, which helps them to stay on task and submit the assignment before the due date. Some teachers allow students to listen to music, which students request and seem to thoroughly enjoy. Teachers also try to help students focus by utilizing the physical space in the classroom; they move the desks around and allow students to spread out. This is interesting, as technology often leads teachers to rearrange the furniture in their class toward a style fit for group learning (Mitchell, 2004). The sample that I observed, however, used a row design for the most part (see Figure 1). Students find incentive to complete assignments by viewing their grade faster through online grading. Because students are continually monitoring their online grades, teachers feel that this method of grading leads to more accuracy when it comes to assessing students. There are numerous creative methods that teachers might employ to help students avoid distraction.

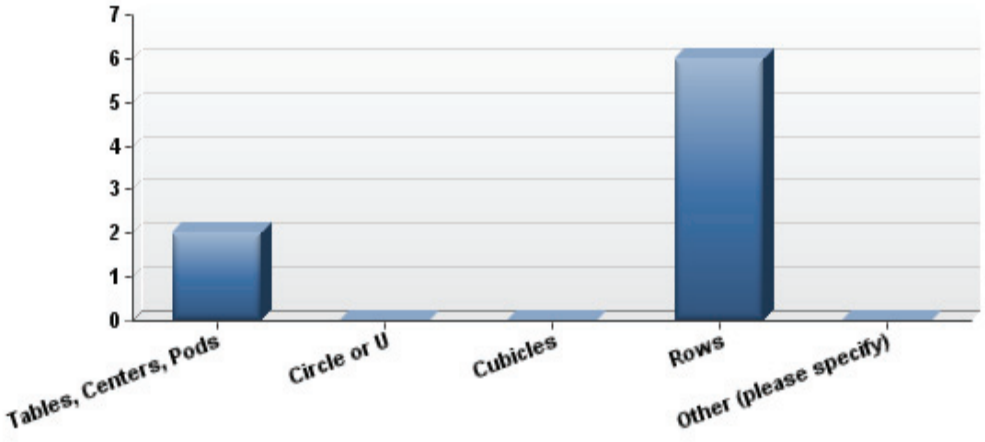

Figure 1. Design of classroom layout (n=8).

How Laptops Help Students Avoid Distractions

When asked if using the laptops has helped their students learn to deal with distraction in general, teachers generally agree that the laptop helps with distraction in multiple, often indirect, ways. One teacher asserts that the block schedule (which includes 90-minute classes) makes for a lengthy class

period in which students might become more easily distracted from the lesson. By having the laptop in class, this teacher notes that students seem more motivated and more interested in the classroom content. Another teacher believes the laptops have taught students "how to multitask and still be effective." Teachers believe the skill of multitasking is one that needs more attention, particularly for males (Wilson, 2005).

Another key way laptops help students avoid distractions is by the types of assignments students may complete using the tool: for example, strict deadlines and peer accountability provide strategic ways for teachers to ensure that students divide their focus in a wise manner. Through tools like discussion boards, Angel makes these strategies easy for teachers to utilize. In this sample, the primary activity for which technology is used is communication. By labeling the activity "communication," the activity could include document preparation, email, presentation, or web development. Fifty percent of teachers and 75% of students use technology to communicate (see Figure 2). The primary way students communicate is through document preparation or e-mail. The next most popular activity for which technology is used is summative assessment. This is illustrated through students having to submit their assignments either through Angel or e-mail. Thirty-eight percent of teachers and 63% of students use the technology primarily for summative assessment (see Figure 2). Finally, the third most popular activity for which technology is used is project-based activities; 25% of teachers and 13% of students use the technology in this way (see Figure 2). Interview data suggests that having students submit assignments through Angel can aid student distraction management.

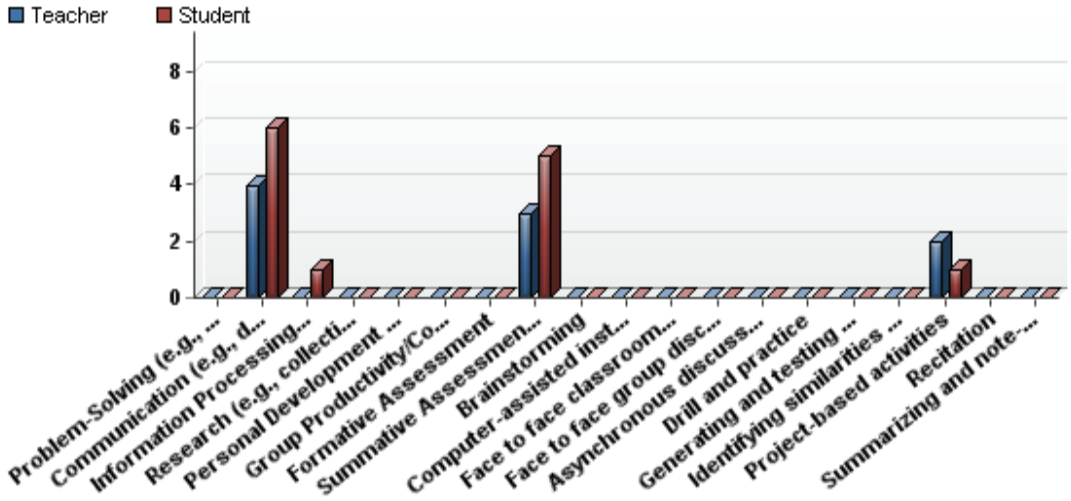

Figure 2 Activity for which technology is used during observations (n=8).

One teacher defines teaching as "a conversation about thinking. And I think it you use the discussion forum…it does open up the possibilities for that." Another teacher comments on the varied activities, such as the quiz or chat feature, that Angel permits. She believes that using multiple features like this helps to keep students guessing (and thus more attentive) about what class activity will occur. By utilizing

the captivating power of technology in class, teachers find that they can pull students' attention toward the lesson and away from distractions.

Online Temptations and the Need to Monitor

All teachers agree that the Internet tempts students to do something non-instructional during class like check personal e-mail or social networking sites. One teacher notes that a student in another class was recently disciplined for attempting to log on to Facebook 12 times in one class period. The school has filters that block sites such as Facebook, but students often use proxies to gain access to their favorite personal websites. Teachers note that when using computers in class, they notice students minimizing windows quickly or trying to hide what they are doing when the teacher walks by. The solution to this problem for most teachers is to keep students so busy that they do not have time to check personal e-mail or social networking sites. Teachers note that "students should be minimizing windows relevant to the course, such as Angel or online research." One teacher likens minimizing windows to his or her own experience: "That's just like when we were in school and people would try to hide their cheat sheets." Teachers understand that students are frequently tempted by the Internet and should be monitored.

Teachers generally feel that most students stay within the parameters of the lesson when using the laptop, whereas there are often a few who like to visit other sites. One teacher likes to have a time in class when laptops are closed and all eyes are on the teacher. This teacher notes that in order to monitor well, "You have to be a vulture. You have to swoop down and you have to make sure that when you say the laptop is down, it's down." This teacher also teaches upper and lower-level honors courses and finds that distraction affects each course in a different way: "It seems to happen more for me with my seniors than with my ninth graders and tenth graders and it may be that ninth graders are fearful that the teacher will catch them; [for] seniors, maybe it doesn't matter so much." This coincides with information from student interviews, which suggests that seniors might feel the need to work on other assignments or check personal e-mail for responses from prospective colleges. When teaching small classes with 10-15 students, teachers have no trouble walking around and monitoring what they are doing; when the class size increases to 30 students, however, teachers have more difficulty monitoring. Teachers note that one method of punishment is confiscating the laptop for 24 hours, but fortunately they do not have to do this very often. It is important for teachers to learn monitoring techniques to help students avoid online distractions in class.

Advantages of 1:1 Learning

Teachers believe that learning to deal with distractions in school is important and will help students in their future endeavors. Teachers see students learning to be responsible, as they are the sole keepers and protectors of their laptops; this develops a sense of ownership for the student. All of the sites students visit online can be checked by the school technician; this fact makes them think twice about viewing inappropriate websites or making unsuitable decisions online. Another key part of being responsible users of their laptops involves bringing laptops to class fully charged, particularly for morning classes. Further, because assignments may be submitted electronically, there is a "no excuses" mentality among students (Corn, Tagsold, &

Patel, 2011). One teacher notes that e-mail has increased his ability to communicate with students "human to human." This teacher can communicate something personal through email, such as a concern over missing work or encouragement during a tough time. Teaching students responsibility is just one advantage of participating in a 1:1 initiative.
Teachers particularly want students to learn that there is a time to check personal e-mail and Facebook and a time to focus on class activities. Teachers assert that students need all the time management training they can get, and they believe that illustrating how to manage their digital lives helps with this skill. Learning these lessons while in a 1:1 environment will help students as they compete in a global economy in the future.

While it can be discouraging when teachers must spend class time managing students' online behavior, one teacher notes, "There was always a slick student before technology that you would have to monitor and pay close attention to." Another teacher expresses sentiments about distractions caused by technology: "Distractions…there will always be distractions. We try to impress upon the kids how important it is to just be responsible for your actions. [We tell them] 'it's your education!'" The table below designates key themes derived from findings.

Table 1
Summary of teacher sentiments by topic.

Topic	**Teacher Sentiment**
How Teachers Help Students Avoid Distractions	Create lessons and assignments that are so engaging that students do not desire any outside stimulation. Keep students so busy that they do not have time to seek distractions online (deadlines, teach from bell to bell).Walk around and monitor. Collect assignments through online drop box. Allow students to move around classroom to work or listen to their preferred music while doing individual assignments. Grade online (more incentive). Emphasize that technology is a tool that can take care of lower-level thinking tasks to free up time for higher-level thinking.
How Laptops Help Students Avoid Distractions	Keep students interested in lesson for 90-minute classes. Teaches how to effectively multitask. Use the Angel discussion boards (strict deadlines, peer accountability).
Online Temptations and the Need	All teachers agree that the Internet tempts students. Seniors try to hide windows more frequently than lower grade levels. Small classes allow for easier monitoring.

to Monitor	
Advantages of 1:1 Learning	Laptops are TOOLS that can greatly enhance learning opportunities as long as students have basic foundational knowledge as well. Laptops teach responsibility. Teachers want students to learn online time management skills.

DISCUSSION

The discussion highlights five themes that emerged from the data: (a) Laptops make learning more fun, (b) Students are less distracted when assignments are challenging, (c) Students are likely more distracted in class than teachers think they are, (d) Teachers and students are developing ways to manage distraction, and (e) Teachers and students understand that technology is here to stay.

Manage Distraction by Incorporating Projects that Students Enjoy

All teachers note the engaging classroom possibilities inherent with 1:1 learning. For teachers with the block schedule, ninety minutes can be a long time for students to listen to a lecture/class discussion, complete a worksheet, or take a test without seeking online distractions. One teacher notes, "I believe [the laptop] has been used as a tool to sort of help what we do, sort of break up boredom with just the regular test." Teachers assert that laptops help students stay engaged for a longer period of time during the 90-minute class. Teachers also find that assessing in multiple ways, such as Glogsters, PowerPoint presentations, or videos, works wonders for making learning more fun for students (Corn, Tagsold, & Patel, 2011). This type of involvement and sharing excites teachers because students perform better when they have a choice in types of assignments they may submit.

Other research notes that the lack of connection between testing and technology is a major problem for teachers who would love to make every class innovative and fun (Corn, Huff, Halstead, & Patel, 2011). Classes assessed by standardized tests are often perceived as more

important than those that are not assessed with such tests (Siskin, 2003). Schoen & Fusarelli (2008) discuss how standardized assessment conflicts with the teaching and learning practices embedded in the 21st century skills movement. For example, the 21st century skills movement focuses more on collaborative, interdisciplinary authentic activities rather than isolated learning and assessment (Schoen & Fusarelli, 2008). The English IV teachers in this study have more free reign with technology than other teachers because their students are not tested through a standardized assessment.

Excellent tools such as FIZZ have been developed based on this concept of providing content to the student outside of class and providing discussion and the opportunity to dive deeper into material during class (Barnhill, 2009). Pecansky-Brock (2011) finds VoiceThread, an online tool which transforms media into a collaborative space through video, voice, and text commenting features, helpful for aligning classes to 21st century learning. These ideas are excellent ways to maintain students' attention. Because students code certain activities as hot (such as social learning in which students connect) and some as cold (being lectured to, completing

dull assignments that are simply "busy work") (see Metcalfe & Mischel, 1999), tools such as FIZZ and VoiceThread are useful for helping students avoid distraction in 1:1 classrooms. Twenty-first century learners have myriad experiences using technology, and teachers would be wise to take advantage of the cultural technology knowledge base shared by many students (what Labbo & Place call *technology funds of knowledge*) (2010).

Manage Distraction by Incorporating Challenging Assignments

Teachers in this sample enjoy creating assignments which challenge students and lead them into higher level thinking activities than a simple Google search could provide. One teacher states, "I want to challenge them... it should be frustrating a little bit. It should cause a little problem. It should encourage them to think." The literature supports this desire for challenging work. Not only does challenging works benefit students academically; it also benefits them by keeping them focused for longer periods of time (Donham, 2011). Relatedly, teachers emphasize the need to remember that the laptop is a tool, and the potential of the tool is only realized when students and teachers desire to learn and work hard to demonstrate their mastery of concepts.

Manage Distraction by Promoting Self-Regulation, Self-Discipline

Fried (2008) found that students who used laptops in a traditional lecture-style university class spent considerable time multitasking and received lower scores on recall tests taken after the lecture. Fried notes that laptops certainly have their place in a classroom specifically designed for their use, but the "unstructured use of laptops in lecture courses is a disadvantage" (2008, p. 912). Fried's research supports the notion that when students pay more attention to the Internet than the lecture, assessment scores decrease; furthermore, this study's outcomes reflect the multitasking myth.

The Multitasking Myth

A traditional teacher believes the laptops have taught students "how to multitask and still be effective." In *Brain Rules for Baby*, Medina (2010) asserts that the best predictor of academic success is not IQ, but self-control. The human brain chooses relevant stimuli from other options, and executive function allows the brain to stay on task and avoid unproductive distractions. Students from this sample would agree with Medina; they feel that reminding themselves of future goals and considering the rewards and punishments involved with seeking outside distractions is the key to staying on task during class.

Ever since the first human found himself learning to survive on this earth, human beings have experienced the double-edged sword that is distraction. When that early human, for example, was reverently watching a beautiful sunset, he would not survive if he did not also hear a hungry lion approaching. Today, this might look like a student who is putting together a beautiful PowerPoint presentation in class who loses his focus when a peer laughs loudly in the hallway. His attention shifts for a moment, and he must bring it back. We give continuous partial attention all the time, and it can be difficult to focus on one task for a long time (Jackson, 2008). In this sample, most students treat online distractions as rewards rather than as the loud laughter in the hallway that might distract a student for just a few seconds. Hembrooke & Gay (2003) find that students who keep their laptops closed during class have higher test

scores on content taught during that class. They note, however, that the key factor for those who keep their laptops open during class is the length of browsing time.

There are different types of multitasking; one type involves doing more than one thing at exactly the same time, such as driving and listening to the radio, or cooking dinner while talking to one's spouse. The other type occurs when people rapidly change from one task to another; an example of this would be found in someone writing an essay for one class and then quickly reading Facebook messages. The first type is called parallel processing, and the second type is called task-switching (Gasser & Palfrey, 2009). Gasser & Palfrey (2009) find that the first type of multitasking (parallel processing) may increase efficiency, while the second type (task-switching) "can decrease efficiency, especially if those tasks demand more challenging cognitive processes" (p. 17).
Students and teachers must understand (and many already do) that parallel processing is a great skill to learn in school (e.g., let a file download while reading the day's assignment). Students should be advised that task-switching, however, can be hazardous to their mental health (Fried, 2008; Kraushaar & Novak, 2010). Teachers in the sample already do a great job of helping students to avoid task-switching through lessons that involve challenging assignments, strict deadlines, and engaging material that keeps students so engaged that they do not have enough time to become bored and seek online distractions. School administrators are also wise to have filters that block websites that typically distract students, such as Facebook and Twitter.

Another idea that emerges from the analysis is that of teaching students not only information and literacy skills, but also how to self-regulate when it comes to online distractions. Teachers report that altering the assessment style from paper and pencil tests to ExamView, an electronic testing system, helped one student because the software presented her with one question at a time, thus helping her focus better (Corn, 2009). Differentiated assessment practices may become the norm as their advantages become apparent over time.

Carr (2010) asserts that the activity of allowing one's mind to focus on irrelevant information rather than that which is meaningful and relevant signals "a reversal of the early trajectory of civilization: We are evolving from cultivators of personal knowledge into hunters and gatherers in the electronic data forest" (p. 2). Researchers, educators, and policy makers must consider what exactly is lost when students are not taught skills to help them manage online distractions.

Teachers are Developing Ways to Manage Distraction

This study reveals numerous methods that teachers use to manage distraction. I received intriguing responses during interviews, and the methods teachers use are as nuanced as the individuals themselves. The literature cites various methods used to help students avoid distraction in high-tech learning environments (Johnson, 2010), including: classroom response systems (clickers) and music (Cole, 2010; Johnson, 2010); setting ground rules at the beginning of the semester for wireless use, including a once-per-class period "no laptop time" (*Wireless in the Classroom: Advice for Faculty*, 2011); and using monitoring software which involves a range of benefits and drawbacks (Corn, 2009; Robinson, Brown, & Green, 2007). Teachers in this sample made use of each method except for

classroom response systems and monitoring software (although the sample school does utilize school wide filters for certain websites).

Manage Distraction by Illustrating the Relevance of Technology

Teachers recognize that the world is much flatter now that most of its inhabitants are connected through the World Wide Web (Friedman, 2005). Glimps (2008) recognizes the need for American schools to better prepare children with physical and health disabilities for a globalized workplace. She points out the need for students to learn about a global world. This may be done particularly through social studies courses, which have traditionally taught a blend of disciplines, including geography, civics, history, and anthropology. Glimps recommends adding more comparative religion and foreign language courses to the curriculum to aid with the 21st century concern of globalization, which aligns with two NETS*S components: Communication/ Collaboration and Digital Citizenship.

This study illustrates how 1:1 learning can help students prepare for the myriad distractions that are available to them in college and beyond through five major themes (see Table 2). Data supports the first theme, "Manage distraction by incorporating project that students enjoy," through reminders to make learning hands-on and reflective of the 21st century skills. Students enjoy it when teachers utilize new technologies for class, even if the teacher needs help from students to make the technology work correctly.

Table 2
Themes and lessons learned.

Theme	**Lessons Learned**
Manage Distraction by Incorporating Projects that Students Enjoy	Make learning hands-on; utilize new technologies for class. By personalizing assignments, students instantly become more interested and engaged.
Manage Distraction by Incorporating Challenging Assignments	Use students' love of socializing to your advantage; have them collaborate on documents or complete problem-solving activities as a group.
Manage Distraction by Promoting Self-Regulation, Self-Discipline	While filters and monitoring software can aid distraction, the best way to keep your students focused is to engage them. Have an open discussion with students about how you define online distraction; recognize that minimizing windows or viewing outside resources might be aiding student learning rather than hindering it. Consider allowing students to listen to music while working on individual assignments and discussing self-regulatory techniques with them, along with why these skills are important for the future.
Teachers are Developing Ways to Manage Distraction	Make work challenging; use deadlines to your advantage; utilize the online grading feature available in most course management systems (such as Angel, Moodle, or Blackboard).
Manage Distraction by Illustrating the Relevance of Technology	Understand that you are teaching during a pivotal moment in history; learning is changing and continuous professional development and the courage to try new teaching methods is imperative.

LIMITATIONS

As the primary instrument of data collection, the researcher had to maintain a constant awareness of her own perceptions and beliefs throughout the research study (Merriam, 1998). The data is self-reported and there is no comparison group; further, due to the small sample size, descriptive data from classroom observations may appear inflated. Because interview data constitutes the primary data source for this study, participants could have responded in a less truthful way fearing punishment. Limitations also include a lack of generalizability due to all participants being from North Carolina high schools.

CONCLUSIONS

As educators look toward the future of 1:1 learning and the many distractions that will be available to students simultaneously, they must learn to utilize data-based distraction avoidance methods in their classrooms. The study helps to move this area of inquiry forward by gaining a thorough understanding of strategies teachers use to manage distraction in 21st century classrooms. By entering a school system that has been provided with high-tech classrooms, informed leadership, and appropriate professional development, this study expands the field of instructional technology and increases the effectiveness of teachers and students so they may excel in school. By employing the techniques that educators say work best for avoiding distractions, teachers can hypothetically prepare students for a lifetime of uninterrupted learning.

REFERENCES

Antonietti, A. & Giorgetti, M. (2006). Teachers' beliefs about learning from multimedia. *Computers in Human Behavior 22*, 267–282. doi:10.1016/j.chb.2004.06.002

Barnhill, C. (2009). FIZZ: Web 2.0 Comes to the Classroom. Featured story, North Carolina State University.

Baumeister, R. F., & Vohs, K. D. (2004). Self-regulation. In C. Peterson & M. E. P. Seligman (Eds.), *Character strengths and virtues: A handbook and classification* (pp. 499–516). Washington, DC/New York: American Psychological Association/Oxford Press.

Bebell, D. (2005). *Technology promoting student excellence: An investigation of the first year of 1:1 computing in New Hampshire middle schools*. Boston: Boston College. Retrieved from https://www.center.uoregon.edu/ISTE/uploads/NECC2005/KEY_6592566/Bebell_BebellNECC2005_RP.pdf

Bell, B. & Kozlowski, S. *(*2002*)*. Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology, 55,* 267–307.

Bogdan, R. C. & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theories and methods* (5th ed.) Boston: Allyn and Bacon.

Carr, N. (2010). Author Nicholas Carr: The Web Shatters Focus, Rewires Brains. *Wired Magazine*. Retrieved from http://www.wired.com/magazine/2010/05/ff_nicholas_carr/all/1

Cavanaugh, C., Dawson, K., White, S., Valdes, N., Ritzhaupt, A., & Payne, D. (2007). *Leveraging Laptops: Effective Models for Enhancing Student Achievement. Project Research Report 2006-07.* Florida Center for Instructional Technology. Retrieved from http://etc.usf.edu/laptops4learning/resources/reports/Final_Report.pdf

Churchill, D. (2006). Teachers' private theories and their design of technology-based learning. *British Journal of Educational Technology, 37*(4), 559-576. doi:10.1111/j.1467- 8535.2005.00554.x

Cole, S. (2010). Quit surfing and start "clicking": One professor's effort to combat the problems of teaching the U.S. survey in a large lecture hall. *History Teacher*, *43*(3), 397-410.

Corbin, J. & Strauss, A. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Los Angeles, CA: Sage.

Corn, J. (2009). *Evaluation Report on the Progress of the North Carolina 1:1 Learning Technology Initiative (Year 2)* (NC State Board of Education Report). Raleigh, NC: Friday Institute for Educational Innovation, North Carolina State University.

Corn, J.O., Halstead, E.O., Tingen, J., Townsend, M., & Campbell, B. (2010, December). *North Carolina 1:1 learning initiative evaluation interim report.* (Vance County Schools, Golden Leaf Foundation Report). Raleigh, NC: Friday Institute for Educational Innovation, North Carolina State University.

Corn, J. O., Huff, J.D., Halstead, E.O., & Patel, R.K. (2011, January). *Examining issues critical to a 1:1 learning environment: Student learning outcomes*. *Evaluation report on the progress of the North Carolina 1:1 learning technology initiative (Year 3)* (NC State Board of

Education Report). Raleigh, NC: Friday Institute for Educational Innovation, North Carolina State University.

Corn, J. O., Osborne, J., & Halstead, E. (September 2008). *Initial report: Evaluation study of the progress of the North Carolina 1:1 learning initiative (Year 1)* (NC State Board of Education Report). Raleigh, NC: Friday Institute for Educational Innovation, North Carolina State University.

Corn, J., Tagsold, J., & Patel, R. (2011). The tech-savvy teacher: Instruction in a 1:1 learning environment. *Journal of Educational Research & Practice 1* (1).

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. (2nd ed.). Thousand Oaks, CA: Sage.

Curtis, R., Merry, L., & Walker, M. (2011). *Can you win? Inside a student-led war room!* Poster presented at the annual meeting of the International Society for Technology in Education, Philadelphia, PA.

Dede, C. (2005). Planning for neomillennial learning styles: implications for investments in faculty and technology. In D.Oblinger & J.Oblinger (Eds), Educating the Net generation (pp. 15.1–15.22). Boulder, CO: EDUCAUSE. Retrieved March 31, 2006, from http://www.educause.edu/educatingthenetgen

Denzin, N.K. and Lincoln, Y.S. (2000), "Introduction: the discipline and practice of qualitative research", in Denzin, N.K. and Lincoln, Y.S. (Eds), *Handbook of Qualitative Research*, Sage, London, pp. 1-28.

Donham, J. (2011). Assignments worth doing. *School Library Monthly, 28* (2): 5-7.

Donnalley, J. (2011). *Practical DyKnow: 10 easy ways to engage students*. Session presented at the annual meeting of the International Society for Technology in Education, Philadelphia, PA.

Ertmer, P. A. Addison, P., Lane, M., Ross, E., & Woods, D. (2000). Examining teachers' beliefs about the role of technology in the elementary classroom. *Journal of Research on Computing in Education, 32*, 54-72. Retrieved from: http://www.iste.org/content/navigationmenu/publications/jrte/issues/volume_321/number_1_fall_1999/examining_teachers_beliefs_about_the_role_of_technology_in_the_elementary_classroom.htm

Fairman, J. (2004). *Trading roles: Teachers and students learn with technology.* Maine Learning Technology Initiative (Research Report #3). Maine Education Policy Research Institute, The University of Maine Office, Gorham, Maine. Retrieved from http://usm.maine.edu/cepare/Reports/MLTI_Report3.pdf

Fraser, M. (2002). Ergonomics for grade school students using laptop computers. Proceedings of The XVI Annual International Occupational Ergonomics and Safety Conference, Toronto. Available online at: http://education.umn.edu/kls/ecee/pdfs/ErgonomicsForGradeSchoolChildrenUsingLaptops_Fraser.pdf

Fried, C. (2008). In-class laptop use and its effects on student learning, *Computers and Education, 50*, 906–914.

Friedman, T. (2005). *The world is flat*. New York, NY: Farrar, Straus & Giroux.

Gasser, U & Palfrey, J. (2009). Mastering Multitasking. *Educational Leadership*, *66* (6), 14.

Harris, W. J., & Smith, L. (2004). *Laptop use by seventh grade students with disabilities: Perceptions of Special Education teachers.* Maine Learning Technology Initiative. (Research Report #2). Maine Education Policy Research Institute, University of Maine Office. Retrieved from http://usm.maine.edu/cepare/Reports/MLTI_Report2.pdf

Hembrooke, H., & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education*, *15* (1), retrieved April 2, 2005, from http://www.hci.cornell.edu/LabArticles/Multitasking_Hembrooke.pdf.

Hesse-Biber, S. (2010). Qualitative approaches to mixed methods practice. *Qualitative Inquiry, 16* (6), 455-468.

Holcomb, L. (2009). Results & Lessons Learned from1:1 Laptop Initiatives: A Collective Review. *TechTrends*, *53*(6), 49-55.

Jackson, M. (2008). *Distracted*. Amherst, N.Y.: Prometheus Books.

Johnson, D. (2010). Taming the chaos. *Learning and Leading with Technology*, 38(3), 20-23.

Kleiger, A., Ben-Hur, Y. & Bar-Yossef, N. (2010). Integrating laptop computers into classroom: Attitudes, needs, and professional development of science teachers-A case study. *Journal of Science Education Technology 19,*187-198. doi:10.1007/s10956-009-9191-1

Lambert, J. (2002). *Digital storytelling : capturing lives, creating community*. Berkeley, CA : Digital Diner Press.

Lane, D. M. M. (2003). *Early evidence from the field: The Maine learning technology initiative: Impact on students and learning*. (Occasional paper #1). Center for Education Policy, Applied Research, and Evaluation; University of Southern Maine. Retrieved from http://usm.maine.edu/cepare/Reports/MLTI_Impact_Students_Learning.pdf

Leu, D., Leu, D., & Coiro, J. (2004). *Teaching with the Internet*, 4th edition, Norwood, MA: Christopher Gordon.

Looking for Technology Integration (2010). *SERVE Center at UNC-G*. Retrieved from http://www.serve.org/lofti.aspx

Lowther, D., Strahl, D., Inan, F., & Bates, J. (2007). *Freedom to learn program Michigan 2005-2006 evaluation report*. Memphis, TN: Center for Research in Education Policy. Retrieved from http://www.memphis.edu/crep/pdfs/Michigan_Freedom_to_Learn_Laptop_Program.pdf
Maninger, R. M. & Holden, M. E. (2009). Put the textbooks away: Preparation and support for a middle school one-to-one laptop initiative. *American Secondary Education, 38*(1), 5-33. Retrieved from http://www.ashland.edu/academics/education/ase/index.html

Mann, D. (2008). *Documenting outcomes from Henrico County Public School's laptop computing initiative*. Retrieved from http://jcots.state.va.us/2008 Content/Materials/Open Ed/Henrico Laptop Report.pdf

Medina, J. (2008). *Brain rules.* Seattle, WA: Pear Press.

Medina, J. (2010). *Brain rules for baby*. Seattle, WA: Pear Press.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3-19. doi:10.1037/0033-295X.106.1.3

Miles, M.B. *&* Huberman, A.M. *(*1994*).* Qualitative data analysis: An expanded sourcebook *(*2nd ed.*)* Thousand Oaks, CA*: Sage Publications.*

Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology, 21*(2), 204-218. doi:10.1037/h0032198

Mischel, W., Shoda, Y., & Peake, P. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*(4), 687-696.

Mitchell Institute (2004). *One-to-One Laptops in a High School Environment. Piscataquis Community High School Study Final Report*. Retrieved from http://www.projectred.org/uploads/Piscataquis_One-to-One_Laptops_Report.pdf

Mollette, M.J., Overbay, A., Corn, J., Townsend, M., & Townsend, L. (2011). *IMPACT Continuation – Annual Report (2009/2010).* Raleigh, NC: Friday Institute for Educational Innovation, North Carolina State University

Oliver, E, Markland, D, & Hardy, J. (2010). Interpretation of self-talk and post-lecture affective states of higher education students: A self-determination theory perspective. *British Journal of Educational Psychology 80* (2): 307–323.

Pecansky-Brock, M. (2011). How and why to "Flip your Classroom" with VoiceThread. Webinar. Retrieved from http://www.teachingwithoutwalls.com/2010/02/voicethread-effective-design-strategies.html

Penuel, W. R. (2006). Implementation and effects of one-to-one computing initiatives: A research synthesis. *Journal of Research on Technology in Education, 38*, 329-348.

Posner, M. & Boies, S. (1971). Components of attention. *Psychological Review, 78*, 391-408.

Posner, M.I. and Rothbart, M.K. (2007) Research on attention networks as a model for the integration of psychological science. *Annu. Rev. Psychol. 58*, 1–23.

Robinson, L. K., Brown, A., & Green, T. (2007). The Threat of Security: Hindering Technology Integration in the Classroom. *Learning & Leading with Technology, 35*(2), 18-23.

Schoen, L., & Fusarelli, L. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st-century schools in an era of accountability. *Educational Policy*, *22*(1), 181-203.
Shapley, K., Sheehan, D., Maloney, C., & Caranikas-Walker, F. (2008). *Evaluation of the Texas technology immersion pilot: Outcomes for the third year (2006-07)*. Retrieved from http://www.etxtip.info/

Silvernail, D. L.. & Lane, D. M. M. (2004). *The impact of Maine's one-to-one laptop program on middle school teachers and students. Phase one summary evidence.* (Research Report #1). Maine Education

(Research Report #1). Maine Education Policy Research Institute, University of Southern Maine Office. Retrieved from http://usm.maine.edu/cepare/Reports/MLTI_Report1.pdf

Siskin, L. S. (2003). Outside the core: Accountability in tested and untested subjects. In Flinders, D.J. & Thornton, S.J. (Eds.). *The Curriculum Studies Readers* (3rd Edition). New York, NY: Routledge.

Vohs, K. D., & Baumeister, R. F (Eds.). (2004). *Handbook of self-regulation: Research, theory, and applications*. New York, NY: Guilford Press.

Warschauer, M. (2006). Going one-to-one: The experiences of cutting-edge schools suggest the whys, the why nots, and the hows of laptop learning programs. *Learning in the Digital Age 63*(*4*) 34-38. Retrieved from http://pdonline.ascd.org/pd_online/tis_tb/el200512_warschauer.html

Weston, M.E. & Bain, A. (2010). The end of techno-critique: The naked truth about 1:1 laptop initiatives and educational change. *Journal of Technology, Learning, and Assessment, 9*(6). Retrieved from http://escholarship.bc.edu/jtla/vol9/6/

Whitcraft, T. & Ibanga, I. (2007). Give 1 Get 1: Closing the Digital Divide. Retrieved from http://abcnews.go.com/GMA/TenWays/story?id=3640025&page=1

Wilson, G. (2005). Abuse of technology can reduce UK workers' intelligence: *HP calls for more appropriate use of always-on technology to improve productivity*. UK. Hewlett Packard study

Wireless in the Classroom: Advice for Faculty. (2011). Computing at UW-Madison Division of Information Technology. http://www.doit.wisc.edu/network/wireless/advice_fac.asp

Wireless in the Classroom: Advice for Students. (2011). Computing at UW-Madison Division of Information Technology. http://www.doit.wisc.edu/network/wireless/advice_stu.asp

Yin, R. K. (2003). *Case study research: Design and methods* (Vol. 5). Thousand Oaks, CA: Sage.