Improving Student Evaluation of Teaching: Determining Multiple Perspectives within a Course for Future Math Educators

Susan Ramlo, sramlo@uakron.edu, The University of Akron

Instructors in higher education are very familiar with the Likert scale Students' Evaluation of Teaching (SET) used to evaluate teaching. Researchers have raised concerns about biases affecting the results of SET surveys, as well as their validity and reliability and use in high-stakes decision making. Here, we demonstrate that Q methodology, identified as an 80 year old mixed method, is better suited to determine the differing student views about a college math education course. We will discuss how Q used for SET can produce results that are more helpful in assisting faculty in improving their teaching effectiveness and in redesigning courses while also providing more informative SET results for faculty evaluations.

Keywords: *Q methodology; student evaluation of teaching; math education; surveys; classroom improvement*

Instructors in higher education and their students are very familiar with the Likert scale Students' Evaluation of Teaching (SET) used to evaluate teaching in college classrooms. Typically, merit, reappointment, tenure, and promotion decisions are made, at least in part, based on the results of such instruments (Baldwin & Blattner, 2003; Kalender, 2015). However, it is commonly known that such evaluations do not necessarily indicate whether or not an instructor was effective in promoting learning in their classroom.

SET instruments typically consist of a set of Likert scale survey questions. Students select responses on this scale, typically from 5 – strongly agree to 1 – strongly disagree, and instructors receive a summary report with the mean values for these responses and possibly an overall mean. The use of a mean value assumes a Gaussian distribution of responses even though it is quite possible that responses are bi-modal or even tri-modal, representing differing views of the classroom experience. Because multiple views of an instructor or course may exist, the researcher selected Q methodology as a means of determining the multiple student views regarding a course for future math educators.

In Q methodology, participants sort statements about a topic. In this study, students sorted 48 statements about different aspects of the course. Many of these statements came from student responses to a questionnaire about what aspects of the course contributed to their success and what did not contribute to their success. In this way, the focus of the study was slightly different from typical SET survey questions because it attempted to focus on students' perceptions about which aspects of the course contributed to their success opposed to a survey that focuses more on students' satisfaction with the course/instructor. Selection of Q methodology was therefore able to provide a more reflective type of SET experience for students that would also provide instructors a more formative, rather than summative, evaluation of their teaching and classroom.

Students' Evaluation of Teaching Surveys

Researchers have raised concerns about biases affecting the results of SET surveys, including the instructor's sex, difficulty of the course, students' grade point average, and other influences not relevant to measuring the teacher's effectiveness. Certainly, various researchers have stated concerns about SET instruments' validity and reliability, even when those instruments are based upon education research (Abrami & d'Apollonia, 1991, 1999; Baldwin & Blattner, 2003; Kalender, 2015; Marsh & Cooper, 1981; Rantanen, 2013). Many SET instruments used in colleges have not been created with that type of theoretical framework and have not been evaluated for reliability and validity. Rantanen (2013) found interrater reliability to be quite low in part because some students are systematically more lenient in their rating whereas others are systematically more severe. In addition, Abrami and d'Apollonia (1999) revealed that some faculty become lenient graders and reduce the quality of their teaching in order to receive better scores on SET instruments. Unfortunately, this response to SET is due to the frequent importance of SET scores for decisions about merit pay, reappointment, tenure, and promotion (Baldwin & Blattner, 2003).

Research on SET spans a century (Marsh et al., 2009). Their perceived purpose is as a summative assessment of the quality of the instruction provided in the class as well as a tool for the instructor to improve their effectiveness in the classroom. Marsh and Dunkin (1997) proposed a reflective SET process and instrument that utilizes common psychometric systems such as investigating validity and reliability, as well as continuous reflection on the instrument through further research. Yet it seems this is certainly the exception in practice rather than the rule. Vanderpol (1959) discussed the same pitfalls of teacher evaluations as those who discussed them several decades later (Marsh & Cooper, 1981; Owens & Williams, 2005). Marsh and Cooper (1981) discovered that students' interest in the subject affected their ensuing Evaluation of Teaching ratings for the courses they evaluated. And although Marsh (1993) later presented support for the consistency of measures related to a specific instrument for SET, it is important to note that consistency (i.e. reliability) does not imply validity. It is generally accepted that although a valid instrument can be considered reliable, reliability does not imply validity (Newman & Newman, 1994). Kalender (2015) found invariability of SET results across instructors across similar courses with different instructors and called into question the validity of SET surveys. Thus, a variety of researchers have questioned the validity of SET instruments for numerous reasons. Given the impact of SET on the careers of academics, noting that SET surveys do not necessarily measure what they purport to measure is an issue that needs to be addressed.

Other issues with SET instruments and analyses also exist. In many ways, Likert scale surveys and their ensuing analyses assume homogeneity of student opinion about their classroom experience. SET results often, for instance, report mean values for each question and frequently also include an overall mean score. The calculation of mean scores assumes a Gaussian distribution of responses (Newman & Newman, 1994). Yet the responses to these questions are infrequently Gaussian. Q methodology does not make such assumptions; to the contrary, Q assumes that multiple perspectives exist within a group and allows a researcher to examine those differing perspectives along with consensus among participants. Earlier investigations by the researcher via informal student interviews and written comments elicited from students support the use of Q methodology to investigate student opinions about their teachers and classroom experiences, as these studies have indicated multiple student opinions within the same classroom. Similarly, the researcher previously demonstrated that students' epistemology about the same course can be represented by multiple views through the use of Q methodology and that such studies reveal more detailed descriptions about these views when compared to Likert scale surveys (Ramlo, 2006/2007, 2008). In other words, Likert scale surveys are not able to reveal the complexity of the opinions within a group like Q methodology (McKeown, 2001). In addition, Likert scale surveys can potentially force compliance with predetermined researcher views (Hilton, Kopera-Frye, & Krave, 2009).

Q Methodology

Q methodology, a mixed research methodology, was specifically created to determine the various opinions in a group on a topic so that those opinions could be described in the type of detail one typically finds in a more qualitative type of analysis (Newman & Ramlo, 2010). William Stephenson specifically developed Q methodology, or Q, as a means of measuring subjectivity (Brown, 1980; Brown, 2008; McKeown & Thomas, 1988; Stephenson, 1953). Q has been used to determine perspectives in a wide variety of fields from marketing research to political science (Brown, 1980; McKeown & Thomas, 1988) but less frequently in education (Brown, 1980). In higher education, the author has used Q methodology to determine views about a variety of situations, from evaluating a newly developed bioinformatics course based on students' views about their learning (Ramlo, McConnell, Duan, & Moore, 2008) to inservice science teachers' views of a professional development workshop (Ramlo, 2012). The strength of Q methodology is that it does not seek to find the mean of a series of responses like typical Likert scale surveys but, instead, provides a way to determine the different views that exist about a topic. Part of the uniqueness of Q is that it correlates (groups) people as opposed to the more typical use of R factor analysis to correlate (group) items (Stephenson & Burt, 1939). The other unique aspect of Q methodology is the Q sort, which allows the participant to be actively engaged in the communication of their view. The Q sort consists of participants placing items, typically statements related to the topic, into a grid, which allows these participants to provide their perspectives about the topic. Using the Q sorts within the factor analyses and the purpose of measuring subjective views puts Q in a different classification from the typical statistical considerations necessary for R method studies (Brown, 1980).

For instance, this study resulted in 14 sorts, which represents nearly all of the class, yet this is a sufficient number of sorts for Q, as it is not suited to large numbers of cases (Brown, 1980; Stainton-Rogers, 1995; Stephenson, 1953). In Q, it is not unusual to be interested in the views of a specific group of people. For example, in this study, the researcher was only interested in the views of one classroom, the only section of the Instructional Techniques of Mathematics course. Thomas and Baas (1993) differentiate *statistical inference*, where the purpose is generalizing to a larger audience from a large, random sample of participants, and *substantive inference*, where the focus is a more qualitative one *about* phenomena. The latter is the case with Q methodology. In Q methodology, the sample size is actually the number of statements, not the number of participants. Thus, the focus is on describing phenomena using a diverse set of statements that capture the range of communications in order to determine and describe the divergent points of view (Brown, 1980). When the sorts are factor analyzed, the Q factors that emerge represent generalizations in that they describe how persons of a certain perspective think about the topic under investigation (Brown, 1980; Thomas & Baas, 1993). The analysis provides a characteristic sort for each factor, which are generalizations based on the responses of several people (Brown, 1980).

The results of Q studies have been found to be replicable (Brown, 1980; Thomas & Baas, 1993). The test-retest reliability of the Q sorts has been shown to be 0.80 or higher (Brown, 1980). In addition, because of the subjectivity of the sorting process, validity is not a concern in Q (Brown, 1980; McKeown & Thomas, 1988; Newman & Ramlo, 2010).

The analyses produce a number of descriptive outputs that are interpreted to confirm or explore people's perspectives. Thus, Q methodology shares many of the focuses of qualitative research while utilizing the type of statistical analyses typically found in quantitative studies (Newman & Ramlo, 2010; Ramlo, 2015; Ramlo & Newman, 2011; Stenner & Stainton-Rogers, 2004). In this way, Q methodology fits into the paradigm of mixed methods research (Newman & Ramlo, 2010; Ramlo, 2015; Ramlo & Newman, 2011; Stenner & Stainton-Rogers, 2004).

As Bazeley (2010) suggests, the integration of qualitative and quantitative research into mixed methods allows the researcher to produce findings that are of greater use and better address the research purpose (Newman, Ridenour, Newman, & DeMarco, 2003). Using Q methodology to group people based on their perspectives, using factor analysis to correlate their views as expressed by their Q sorts, is a more effective method of grouping people than using surface characteristics such as race, sex or academic major because surface characteristics do not necessarily determine similar perspectives (Ramlo & Newman, 2010). Stakeholders often view situations differently and it is important to ascertain their needs to make effective improvements (McNeil, Newman, & Steinhauser, 2005).

The Q Sample

In this study, a course in techniques of instruction for pre-service math teachers was evaluated using Q methodology. The study's Q sample was slightly changed from the Q sample developed earlier for a physics course (Ramlo, 2011). For the physics SET study, the concourse first included various items associated with the course, such as classroom technology (screen-shots with recordings, clickers, etc.), assessments (homework, quizzes, tests), learning activities, and course difficulty. This resulted in 52 statements. The first semester physics students were also surveyed about their experiences in the course. Specifically, each student was asked to write down at least three things that were important (helpful) for their success in the course and three things that were not helpful for their success in the course. Combined, this created a concourse of 103 statements. Forty-eight statements were selected from this concourse for the Q sample such that the statements selected represented the range of communications related to this course.

The 48 statements related to the SET of the physics courses to be evaluated had to be changed slightly for the SET related to the Instructional Techniques of Mathematics (ITM) course examined in this study because of differences in assignments and learning activities (e.g., no laboratory time in the ITM course). The 48 SET statements sorted by the pre-service math teachers are listed in the Appendix.

The Course Evaluated

Instructional Techniques of Mathematics (ITM) is an upper level course for undergraduate mathematics education majors and masters-level students in an AYA[1] (secondary education) certification program. Students take a field experience course concurrently that involves a course placement (where all students and the instructor observe and teach at the same school) and individual placements (each student is placed at his or her own school). Both the field experience and ITM courses are designed to prepare students for student teaching in a subsequent semester. The spring semester class in this study consisted of three graduate students, all male, and 12 undergraduate students (9 females, 4 juniors). All of these students

[1] AYA is an acronym for Adolescent to Young Adult and represents the secondary (high school) education level.

were planning on either completing their student teaching the following fall or the next spring semester. The course description is as follows:

> *This course will discuss traditional and recent trends in mathematics methods and curriculum. Primary topics will address the various philosophies of education and their application to the mathematics classroom, models of math curriculum, models of mathematical instruction, and models of evaluation.*

Sorting & Analyses

Students were asked to first sort each of the 48 statements into one of three piles: "Most unimportant for my learning in this course," "Neutral," and "Most important for my learning in this course." Then the participants sorted these statements into the grid that is shown in Figure 1. Each statement is represented by a number and these numbers were entered into the grid shown. Asking students to sort based upon their learning in the course was selected because this was deemed more important than whether or not they enjoyed different aspects of the course. In addition, the statements selected for sorting reached beyond what the instructor potentially contributed to student learning and success and included student actions and other "outside" influences.

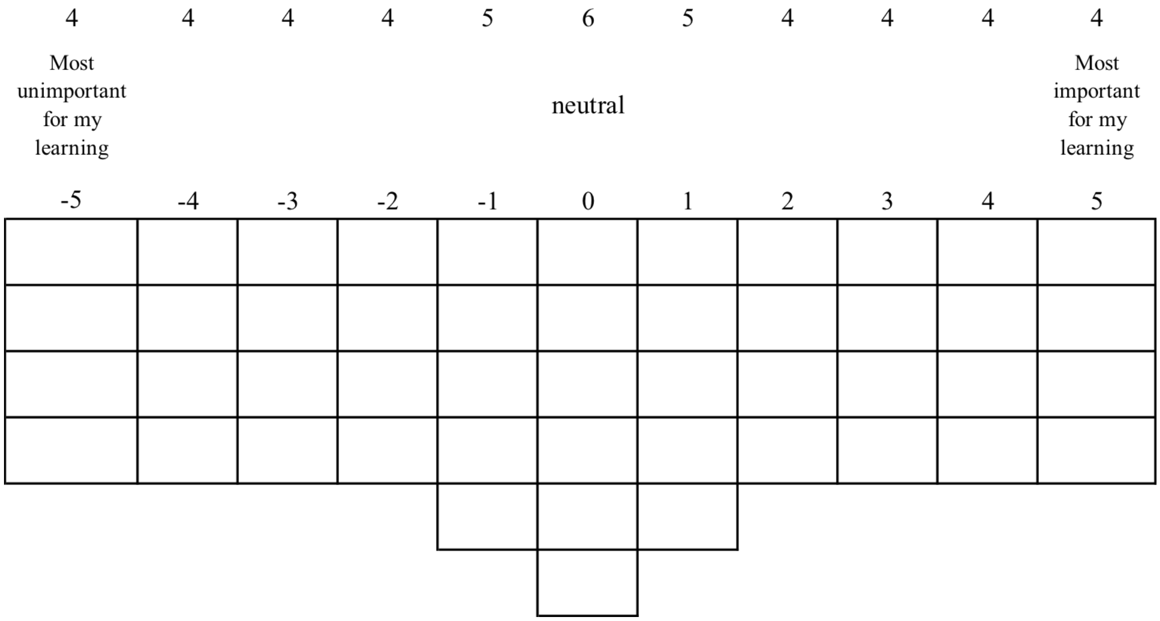

Figure 1.

Sorting grid for this study.

Determining the Factors

The sorts were entered into PQ Method (Schmolck, 2002), a specially designed computer program for factor analyzing and producing the results from the initial Q sorts. Principal components were used to extract two factors and Varimax was used for rotation. Fourteen students out of 15 sorted the 48 statements; one female student was absent the day of the sort. Table 1 contains the two factors for this study with X's indicating defining sorts. Only those sorts/participants who are represented by a factor (X) are used in the remaining analyses. Nine students are represented by Factor 1 and three are represented by Factor 2. The two remaining sorters have loadings that are nearly equal for the two factors.

Table 1.

Factor Matrix With an X Indicating a Defining Sort

QSORT	ID	1	2
1	F_Bach1	0.5572X	0.3922
2	F_Bach2	0.6607X	0.1627
3	F_Bach3	0.8064X	-0.0718
4	F_Bach4	0.6435X	0.3275
5	F_Bach5	0.4606	0.5154
6	M_Bach1	0.6975X	0.2332
7	M_Mast1	0.2462	0.7978X
8	M_Mast2	-0.0614	0.7486X
9	M_Mast3	0.5094	0.5079
10	F_Bach6	0.6005X	0.2919
11	M_Bach2	0.3099	0.6468X
12	M_Bach3	0.7497X	0.2249
13	F_Bach7	0.5118X	0.4686
14	F_Bach8	0.6472X	0.2104

Note. Bach stands for undergraduate level student and Mast stands for master's degree student; Female and male students are identified with an F or an M, respectively.

Importance of Field Experience

Examining the representative sorts for each factor, it becomes apparent that both factors have nearly the same four +5 statements (three out of four are the same). These statements are 37, 2, 28, and 1 for Factor 2 and 37, 2, 26, and 28 for Factor 1. Each of these statements is related to the field experience component of the course. The remaining statements differentiate the two factors and the table of distinguishing statements is the most useful for describing these two factors and therefore the students' views of the other aspects of the course.

Table 2.

Those Statements That Distinguish Factor 1 From Factor 2

Statement No.	Statement	Factor 1 Grid Position	Factor 2 Grid Position
26	Feedback on your teaching within field placement (course &/or individual)	5*	2
44	Feedback on assignments instructor posted via online classroom management system	4*	-2
31	My interest in the topics / math / teaching	3	2
40	Having info posted on online classroom management system (News, readings, assignments, etc)	3*	-1
23	Instructor's knowledge of subject	3*	-2
18	Instructor motivated me to do my best	2*	-4
45	Talking about assessment with field teacher(s)	2*	-4
47	Mock Praxis III Assignment	1*	-5
5	Classroom experiences in Education Building	1*	-3
30	Reflection assignment on assessment	1*	-4
14	Fairness of instructor	0	2
27	Learning to connect math ideas & concepts	-1*	4

25	Instructor's selection of course topics	-1	-3
33	My need to get a good grade this course	-1*	4
8	Manipulatives group project & presentation	-2*	4
24	Instructor's availability (in person, online, email, etc)	-2	1
36	Presentation on developmental math at the University of Akron	-2*	3
4	Observations at class field experience - Early College	-3*	4
12	Math Concept Demonstration	-4	-1
39	Power Algebra presentation by local school district representative	-4*	3
34	My other commitments (job, life, family)	-4*	1
15	Fear of failing	-5*	-2

Note. $P < .05$; asterisk (*) indicates significance at $P < .01$

Describing the Two Factors

Table 2 contains those statements that distinguish Factor 1 from Factor 2. From this table, we see that those represented by Factor 1 believed that communication within the course was important for their learning, whether it was feedback from the instructor on assignments or on the online course management system (Springboard) or from talking with their field placement instructors about assessment. Unlike Factor 1, the Factor 2 view did not seem to focus on communication or the instructor's knowledge of the topic or motivation. For the Factor 2 view, getting a good grade was important for their learning (Statement 33 at +4 opposed to -1 for Factor 1). These students felt learning to connect math ideas and concepts was important

(Statement 27 at +4 opposed to at -1). Yet these students placed the Math Concept Demo statement at -1 (-4 for Factor 1) and the Praxis III statement at -5 (+1 for Factor 1). Student feedback indicated that students felt that these assignments were more "busy work" than a learning experience. Perhaps this is why the Factor 2 view also indicated that the Manipulatives Group Project & Presentation was important for their learning (Statement 8 at +4 opposed to -2 for Factor 1).

Instead of simply taking information from a table, like Table 2, we can also summarize the important statements for each of these factors using an image. Pictorial representation of data can organize and summarize research data in a way that tabular versions of the same data cannot (Dickinson, 2010). Word clouds are a method of visually representing text data including the findings from Q studies (Ramlo, 2011). Most typically, word clouds are used to present keywords from websites in a visual display keywords in websites (Cui et al., 2010). Free word cloud creation is available at websites such as www.wordle.net. The more prominent (larger text size) the word is in the word cloud, the more frequently it appeared in the text provided.

Figure 2.
Word cloud for Factor 1's +5, +4 and +3 statements that distinguished it from Factor 2.

Figures 2 and 3 are word clouds made for Factors 1 and 2 respectively, using the +4 and +3 statements for each that were distinguishing statements for the factor. Figure 2 enforces the idea that Factor 1 is very focused on communication (such as items posted on Springboard including instructor feedback, news announcements, and reading materials). Alternatively, Figure 3 seems more complex with more words included in this word cloud. This figure reinforces the earlier mentioned findings for the Factor 2 view but we can also see that the word "presentation" is the most dominant (font size) in this word cloud. The word "presentation" is found in three of the distinguishing statements for Factor 2 with grid positions of +4 and +3. Two of these involve the students as observers of a presentation and the third involves students presenting. In each case the presentations were visual and hands-on. They were also less abstract than some of the other communications and perhaps this differentiates these presentations from the other communications that the Factor 1 view found to be so important to their learning but the Factor 2 view considered less important or unimportant. Recall too that both factors found that involvement in the field experience was important, at the +5 grid position, for their learning.

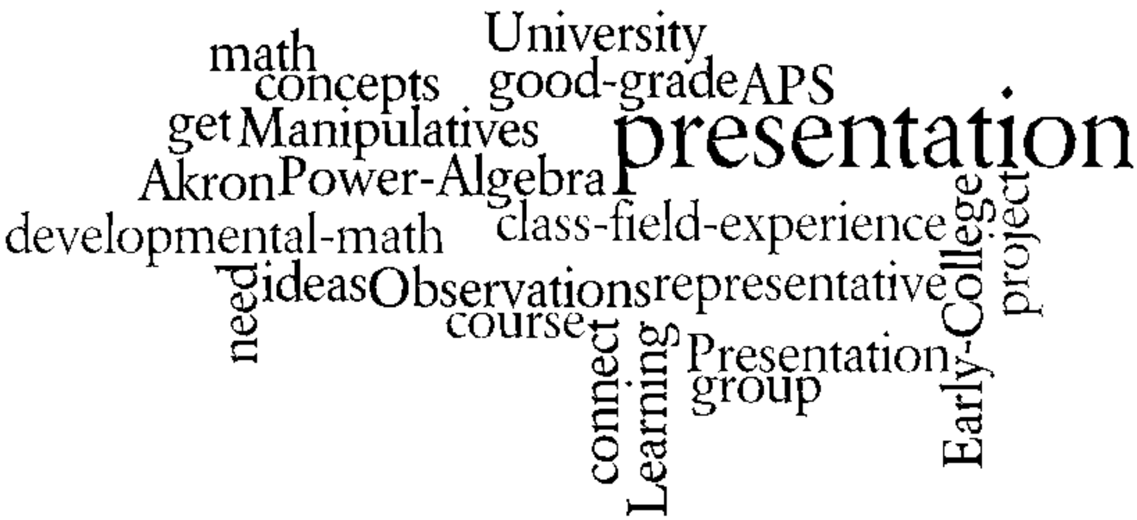

Figure 3.

Word cloud for Factor 2's +4 and +3 statements that distinguished it from Factor 1.

Consensus

The Q methodology analyses also include consensus statements that represent the consensus of the two views that emerged in the study. In this case, the consensus indicates what aspects of the course the two factors/views agreed on. We already discussed the agreement that the field placements were the most important factor for student learning in this set of courses. The other aspects of the courses that these two views agreed on is also important. Table 3 contains the 26 statements that represent consensus at the P < .05 level.

In a number of ways, Table 3 is the most useful for analyzing the different aspects of the course and how they relate to student learning, as seen through their eyes. This table has been sorted not by statement number, as it appears in the PQ Method printout, but by Factor 1 z-score such that the order of the consensus statements begins with those both factors felt as most important for their learning and continues to those they felt were most unimportant for their learning.

Table 3.

Those Statements That Represent Consensus Among Factors 1 and 2

Statement No.	Statement	Factor 1 Grid Position	Factor 1 z-score	Factor 2 Grid Position	Factor 2 z-score
37	Course field experience at East High School - teaching	5	1.97	5	1.77
2	Observations at class field experience - East High School	5	1.79	5	2.05
28	Teaching at individual field experience	5	1.7	5	1.82
1	Observations at individual field experience	4	1.53	5	1.55
48	Mock interview at Career Center	4	1.39	3	1.13
16	Lesson plan creation (individual &/or in-class field)	4	1.18	3	0.99
17	Feedback on teaching in field placement (individual)	3	1.07	2	0.54
20	Instructor's expectations for student's work	2	0.48	1	0.14
29	My desire to learn in this course	2	0.47	0	-0.09
22	Instructor's helpfulness	1	0.42	0	0.04

19	Instructor's enthusiasm for teaching	1	0.32	1	0.22
41	Career Center presentation to our class	0	-0.01	0	-0.06
6	Previous education courses taken	0	-0.02	1	0.28
35	My own perseverance / not giving up	0	-0.05	-1	-0.26
11	Small group exercises in class (e.g. making concept questions)	0	-0.09	0	-0.09
42	The connections between this course's topics and my major	0	-0.12	0	-0.13
32	My natural ability related to the course topics/material	-1	-0.28	-1	-0.35
7	Reflection on Education Policy	-1	-0.49	-3	-1
3	Clicker questions during lectures in class	-2	-0.58	0	-0.04
9	Reflection on math standards	-3	-0.9	-5	-1.45
13	Readings on Springboard (like Alfie Kohn & Swimming Lessons)	-3	-0.94	-4	-1
10	NCTM Standards-based strategy portfolio	-3	-0.98	-1	-0.45
43	The pace of the course (rate at which material was covered)	-4	-1.3	-5	-1.46
21	Previous math courses taken	-5	-1.32	-3	-0.9

38	Reading "The World is Flat"	-5	-1.37	-2	-0.87
46	Access to recorded lectures (audio & screen via Panopto)	-5	-1.93	-5	-1.96

Note. All listed statements are non-significant at P > .05.

In Table 3, we again see the importance of the field experience for both factors. Other aspects of the field experience such as making lesson plans was important for both factors (+4, +3), as was feedback on their teaching at their individual field experience positions (+3, +2). We can also see that the mock interview at the career center was an important assignment for these pre-service teachers even though it did not directly pertain to mathematics instruction. However, the earlier presentation by the career center was at the neutral "0" position for both factors. Aspects of the course that students felt were unimportant for their learning were more passive than those they felt were important. For instance, the assigned readings were deemed unimportant (statements 13 and 38) as were the classroom screen-capture recordings and the clickers used in class as a polling and formative assessment tool. Previous math course work and assignments that were meant to help the pre-service teachers reflect on their practice within the context of education policy and math education standards also were deemed unimportant in their opinion.

Discussion of Results

Analyses of the 14 Q sorts led to two factors/views about the course. Although distinct, a great deal of consensus exists between the two factors. Both factors/views agree that the field experience (observations and teaching) was extremely important to their learning. The observing and teaching aspects of the field experiences were unlike most of these pre-service teachers' educational experiences in college. Other consensuses about the course also represent their focus on course aspects that they felt directly impact their future teaching careers. The pre-service teachers preferred more active, pragmatic course activities (mock interviews at the career center) rather than those that were more passive (presentation by career center staff). They felt that course activities such as reading and reflecting on education policy were unimportant for their learning. Overall, these students preferred activities that they did not label as "busy work" and that were more directly related, in their opinions, to their student teaching and future positions as math teachers.

The two factors are differentiated by looking beyond the field experience statements that were at the +5 position for both factors. Examining the distinguishing statements for the two factors revealed the most pertinent information for examining these differences. Based on the Table 2 data and the resulting word clouds, Factor 2 appears to represent the most pragmatic students for who applying their teaching in the math education class or in a school setting was the most important. They were interested in learning more from teachers (APS Power Algebra and developmental math presentations) as well as presenting themselves (Manipulatives project). Although these students were also more focused on their final grade than the Factor 1 students, Factor 2 was named "***The Practitioners***" because of their focus on real-world types of experiences which concentrate on aspects they felt were directly related to teaching in their future classrooms.

In contrast, Factor 1 was named "***The Thoughtful Communicators***" because of their focus on communication. This view felt that feedback on assignments and their teaching was important for their learning. They sought information whether it was on the online course management system (Springboard which is more broadly known as Desire2Learn) or by talking with field placement teachers. This view, in contrast to the Factor 2 view, felt that the professor's knowledge was important to their learning. Those represented by this view wanted to talk about teaching and reflect on their practice.

Unfortunately, the regular SET results are not available for this course and instructor. However, these two approaches cannot practically be compared as they have very different results. The regular SET summary provides average Likert-scale values for each question in the survey. This type of presentation presumes a singular view of the classroom experience. The Q methodology study here instead groups people based on similar views. This approach assumes multiple views may exist and, in fact, two views emerged in this study.

Conclusions

In this study, a course in techniques of instruction for pre-service high school math teachers (ITM) was evaluated using Q methodology. Various aspects of the field experience (observations and teaching) component of the course was undoubtedly believed by the students to be the most important for their learning, regardless of which Q factor/view represented the pre-service teachers. Overall, the consensus statements represent the overall pragmatic stance of these students within this course that is meant to prepare students for student teaching. For example, the pre-service teachers ranked statements high (most important for their learning) if they felt those items directly influenced their success as teachers (lesson plan development and mock interviews). Alternatively, statements related to more passive activities (specific readings) were placed in the "most unimportant" section of the grid by both factors/views. Examining the distinguishing statements led to the naming of the factors. The **"*Thoughtful Communicators*"** (Factor 1) desired to have two-way communication with their instructor and their field experience teachers and felt the communication had a positive impact on their learning. **"*The***

***Practitioners*"** (Factor 2) were far more pragmatic and preferred hands-on, experiential learning experiences that were directly related to practice.

Thus, this study demonstrates how Q used for student evaluation of teaching (SET) can produce results that are more descriptive of the various student views that exist within the same classroom. Further, issues with psychometrics such as validity and reliability that are often not addressed with SETs are not a concern with Q methodology. Other concerns discussed within the SET literature that are often of concern, such as instructors purposefully making courses easier and good grades attainable to improve SET scores on Likert scales, are not relevant within this study because of its use of Q methodology.

In addition, this study demonstrates that consensus about a course also provides insight about what students feel is the most important and unimportant for their learning. These types of results provide information that faculty may find helpful as they reflect on their practice and classroom activities. In other words, using Q for SET can help instructors improve their teaching effectiveness and redesign courses. This type of information can also provide more informative SET results for those who must review them for decisions associated with merit, reappointment, tenure, and promotion.

This study also has implications for SET use where, with their use of Likert scales and summaries that typically include means for each question, the assumption is a singular student viewpoint. In other words, the practice of using means for each question in a Likert-scale SET survey assumes a homogeneous view of the classroom experience and instructor but the findings in this study demonstrate that student views can be heterogeneous. The use of Q methodology for SET also could allow instructors to include specific statements about what they are doing in their courses in addition to a standard set of statements, to customize the evaluation to what is actually being done in the course rather than depending upon a singular, generic Likert scale survey. Future research should include other SET studies using Q methodology in a variety of courses to better demonstrate its usefulness for SET across multiple disciplines.

References

Abrami, P. C., & d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching Effectiveness—Generalizability of "N = 1" research: Comment on marsh (1991). *Journal of Educational Psychology, 83*(3), 411.

Abrami, P. C., & d'Apollonia, S. (1999). Current concerns are past concerns. *American Psychologist, 54*(7), 519.

Baldwin, T., & Blattner, N. (2003). Guarding against potential bias in student evaluations. *College Teaching, 51*(1), 27.

Bazeley, P. (2010). Computer assisted integration of mixed methods data sources and analyses. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (Second ed., pp. TBD). Thousand Oaks, Calif.: SAGE Publications.

Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. New Haven, CT: Yale University Press.

Brown, S. R. (2008). Q methodology. In L. M. Given (Ed.), *The Sage Encyclopedia of Qualitative Research Methods* (pp. 700-704). Thousand Oaks, CA: SAGE.

Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M., & Qu, H. (2010). Context-preserving, dynamic word cloud visualization. *IEEE Computer Graphics & Applications, 30*(6), 42-53. doi:10.1109/MCG.2010.102

Dickinson, W. B. (2010). Visual displays for mixed methods findings. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (Second ed., pp. 469-504). Thousand Oaks, Calif.: SAGE.

Hilton, J. M., Kopera-Frye, K., & Krave, A. (2009). Successful aging from the perspective of family caregivers. *Family Journal, 17*(1), 39-50.

Kalender, İ. K. (2015). Measurement invariance of student evaluation of teaching across groups defined by course-related variables. *International Online Journal of Educational Sciences*, *7*(4), 69-79.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education, 64*(1), 1-18.

Marsh, H. W., & Cooper, T. L. (1981). Prior subject interest, students' evaluations, and instructional effectiveness. *Multivariate Behavioral Research, 16*(1), 83.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*(3), 439-476. doi:10.1080/10705510903008220

Marsh, H. W., & Roche, L. A. (1997). Making student's evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187.

McKeown, B. (2001). Loss of meaning in Likert scaling: A note on the Q methodological alternative. *Operant Subjectivity, 24*, 201-206.

McKeown, B., & Thomas, D. (1988). *Q methodology*. Newbury Park, CA: Sage Publications.

McNeil, K. A., Newman, I., & Steinhauser, J. (2005). *How to be involved in program evaluation: What every administrator needs to know*. Lanham, Maryland: Scarecrow Education.

Newman, I., & Newman, C. (1994). *Conceptual statistics for beginners* (2nd ed.). Lanham, Md.: University Press of America.

Newman, I., & Ramlo, S. (2010). Using Q methodology and Q factor analysis to facilitate mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (2nd ed., pp. 505-530). Thousand Oaks, CA: Sage.

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P., Jr. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 167-188). Thousand Oaks, CA: Sage.

Owens, D., & Williams, C. (2005). Ongoing student evaluation essential to course improvement. *Distance Education Report, 9*(19), 5-5.

Ramlo, S. (2006/2007). Student views of learning in an introductory college physics course: A study using Q methodology. *Operant Subjectivity, 30*(1 / 2), 52-63. doi:10.15133/j.os.2006.008

Ramlo, S. (2008). Student perspectives on learning physics and their relationship with learning force and motion concepts: A study using Q methodology. *Human Subjectivity, 2*(1), 73-90.

Ramlo, S. (2011). Using word clouds to present Q methodology data and findings. *Human Subjectivity, 9(2)*, pp 99-111. Available at http://www.dbpia.co.kr/Journal/ArticleDetail/1633388.

Ramlo, S. (2012). Inservice science teachers' views of a professional development workshop and their learning of force and motion concepts. *Teaching and Teacher Education, 28(7),* pp 928-935. doi: 10.1016/j.tate.2012.04.002

Ramlo, S. (2015). Theoretical significance in Q methodology: A qualitative approach to a mixed method. *Research in the Schools, 22*(1), pp 68-81.

Ramlo, S., McConnell, D., Duan, Z., & Moore, F. (2008). Evaluating an inquiry-based bioinformatics course using Q methodology. *Journal of Science Education and Technology, 17*(3), 219-225.

Ramlo, S., & Newman, I. (2010). Classifying individuals using Q methodology and Q factor analysis: Applications of two mixed methodologies for program evaluation. *Journal of Research in Education, 21*(2), 20-31.

Ramlo, S., & Newman, I. (2011). Q methodology and its position in the mixed methods continuum. *Operant Subjectivity: The International Journal of Q Methodology, 34* (3), pp 213-214. doi:10.15133/j.os.2010.009

Rantanen, P. (2013). The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*, *38*(2), 224-239.

Schmolck, P. (2002). *PQMethod manual mirror.* Unpublished manuscript. Retrieved April 29, 2004, Retrieved from http://www.rz.unibw-muenchen.de/~p41bsmk/qmethod/.

Stainton-Rogers, R. (1995). Q methodology. In J. A. Smith, R. Harré & L. van Langenhove (Eds.), *Rethinking methods in psychology* (pp. 178-192). London; Thousand Oaks, Calif.: Sage Publications.

Stenner, P., & Stainton-Rogers, R. (2004). Q methodology and qualiquantology: The example of discriminating between emotions. In Z. Todd, B. Nerlich, S. McKeown & D. D. Clarke (Eds.), *Mixing methods in psychology* (pp. 101-120). Hove, NY: Psychology Press.

Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.

Stephenson, W., & Burt, C. (1939). Alternative views on correlations between persons. *Psychometrika, 4*, 269-281.

Thomas, D.D., & Baas, L.R. (1993). The issue of generalization in Q methodology: "Reliable Schematics" revisited. *Operant Subjectivity, 16,* 18-36.

Vanderpol, J. (1959). Student Opinion—Sacred cow or booby trap? *Journal of Teacher Education, 10*(4), 401-412.

Appendix

1. Observations at individual field experience
2. Observations at class field experience - East High School
3. Clicker questions during lectures in class (Zook Hall)
4. Observations at class field experience - Early College
5. Classroom experiences in Zook Hall
6. Previous education courses taken
7. Reflection on Education Policy
8. Manipulatives group project & presentation
9. Reflection on math standards
10. NCTM Standards-based strategy portfolio
11. Small group exercises in class (e.g. making concept questions)
12. Math Concept Demonstration
13. Readings on Springboard (like Alfie Kohn & Swimming Lessons)
14. Fairness of instructor
15. Fear of failing
16. Lesson plan creation (individual &/or in-class field)
17. Feedback on teaching in field placement (individual)
18. Instructor motivated me to do my best
19. Instructor's enthusiasm for teaching
20. Instructor's expectations for student's work
21. Previous math courses taken
22. Instructor's helpfulness
23. Instructor's knowledge of subject
24. Instructor's availability (in person, online, email, etc)
25. Instructor's selection of course topics
26. Feedback on your teaching within field placement (course &/or individual)
27. Learning to connect math ideas & concepts
28. Teaching at individual field experience
29. My desire to learn in this course
30. Reflection assignment on assessment

31. My interest in the topics / math / teaching
32. My natural ability related to the course topics/material
33. My need to get a good grade this course
34. My other commitments (job, life, family)
35. My own perseverance / not giving up
36. Presentation on developmental math at the University of Akron
37. Course field experience at East High School - teaching
38. Reading "The World is Flat"
39. Power Algebra presentation by APS representative
40. Having info posted on Springboard (News, readings, assignments, etc)
41. Career Center presentation to our class (Zook)
42. The connections between this course's topics and my major
43. The pace of the course (rate at which material was covered)
44. Feedback on assignments instructor posted via Springboard
45. Talking about assessment with field teacher(s)
46. Access to recorded lectures (audio & screen via Panopto)
47. Mock Praxis III Assignment
48. Mock interview at Career Center