Kentucky Journal of Excellence in College Teaching and Learning

Editor, Samuel Hinton
Professor, Comparative Education/Educational Studies
Eastern Kentucky University

Managing Editor, Sherwood Thompson
Professor, Educational Leadership
Eastern Kentucky University

Technical Editor, Shawn White
Adjunct Professor, English
Eastern Kentucky University
Board of Reviewers

Lola Aagaard, Morehead State University
Marlisa Austin, Jefferson Community and Technical College
Shelly Chesnut, Grant County Middle School
Roger Cleveland, Eastern Kentucky University
Sandy Conners, Brescia University
Beverly Downing, Kentucky State University
Steve Dwinnells, University of Louisville
Susan Edington, Murray State University
Jody Fernandez, Morehead State University
Barbara Fiehn, Western Kentucky University
Larry Grabau, University of Kentucky
Larry Guin, Murray State University
Gary Hackbarth, Northern Kentucky University
Steve Kerlin, Northern Kentucky University
George Kontos, Western Kentucky University
Gashaw Lake, Kentucky State University
Richard Lewine, University of Louisville
Andy Martin, Kentucky State University
Sonia Michael, Eastern Kentucky University
Chris Miller, Morehead State University
James Obielodan, Kentucky State University
Peggy Pattman-Munke, Murray State University
Kristine Pfendt, Northern Kentucky University
Emily Piercy, Bluegrass Community and Technical College
Norman Powell, Eastern Kentucky University
Narayanan Rajendsan, Kentucky State University
Caroline Reid, Eastern Kentucky University
April Richards, University of Kentucky
Dana Seay, University of Louisville
Melissa Shirley, University of Louisville
Mary Shortridge, Ashland Community and Technical College
Karen Skaff, University of Kentucky
Huifang Sun, Western Kentucky University
Deneia Thomas, Eastern Kentucky University
Shelley Thomas, University of Louisville
Susan Weaver, University of the Cumberlands
Paul Wirtz, Northern Kentucky University
Yuejin Xu, Murray State University

Board of Advisors

Melissa Bell, Kentucky Council for Post-Secondary Education
Jill Compton, Kentucky Council for Post-Secondary Education
Jennifer Miles, Kentucky State University
Enid Wohlstein, Kentucky Virtual Library
Charlie Sweet, Eastern Kentucky University
Marie E. Brown, University of Louisville
Hal Blythe, Eastern Kentucky University
Bill Burke, University of Kentucky
Jerry Daday, Western Kentucky University
Ame Almgquist, Northern Kentucky University
Janet McCoy, Morehead State University
Jamie Justice, Kentucky Community and Technical College System
Sarah Lashley, Centre College
Deborah Castiglione, University Kentucky

ISSN 1943-7943 PRINT
ISSN 1943-7935 ONLINE
Table of Contents

In This Issue
Samuel Hinton, Eastern Kentucky University 8

Faculty Inter-Rater Reliability of a Reflective Journaling Rubric
Mari Alschuler, Youngstown State University 9

Measuring Student Satisfaction in Online Mathematics Courses
Antoinette Davis, Eastern Kentucky University 21

Preservice Teachers’ Perceptions of Mathematics through Drawings
Adam Akerson, Stephen F. Austin State University 38

Course Difficulty and its Association with Student Perceptions of Teaching and Learning
Amanda Joyce, Murray State University 54
Overview

The Kentucky Journal for Excellence in College Teaching and Learning is a peer-reviewed electronic Journal sponsored by the Kentucky Council for Post-Secondary Education Faculty Development Group. General categories for articles include theory, research, and practice. In addition, the Journal will publish manuscripts containing media reviews, profiles, and commentary. The Journal initially targets the professional development of college and university faculty within the Commonwealth. Its primary goal is to enhance student learning by promoting excellence in teaching in higher education institutions in Kentucky and beyond.

Submission Guidelines for Research, Theory, and Practice Papers, respectively

1. Submit original manuscripts as MS Word documents to the Journal website.
2. Use APA style when preparing the manuscript. Please refer to the American Psychological Association Manual of Style, Sixth Edition.
3. Include an abstract of 200 - 250 words for research, theory, and practice.
4. Use one-inch margins and double-space your text.
5. The length of your manuscript should be between 3,000 – 5,000 words in Times New Roman, 12-point font.
6. Use a minimum of tables. Provide image-ready tables, figures, graphs, and charts on separate pages. Include only essential data in the tables, figures, graphs, and charts. Combine tables whenever possible.
7. Do not include the author’s name(s), positions, titles, or places of employment on the cover page to safeguard anonymity.
8. Do not use generic masculine pronouns or other sexist terminology in your manuscript.

Peer-reviewers read and recommend articles for publication under the research, theory, and practice categories. Articles published in the Kentucky Journal of Excellence in College Teaching and Learning, reflect the views of the authors and not of the Editors, Editorial Board, or Eastern Kentucky University.

Submission Guidelines for Media Review, Profile, and Commentary

1. Submit original manuscripts as MS Word documents to the Journal website.
2. Use APA style when preparing the manuscript. Please refer to the American Psychological Association Manual of Style, Sixth Edition.
3. Include an abstract of 120 words.
4. Double-space your manuscript and insert one-inch margins.
5. The length of your manuscript should be 2,000 to 3,000 words in Times New Roman, 12-point font.
6. Use a minimum of tables. Type tables on separate pages. Include only essential data in the tables. Combine tables whenever possible.
7. Do not include author’s name(s), positions, titles, or places of employment on the cover page to safeguard anonymity.
8. Do not use generic masculine pronouns or other sexist terminology in your manuscript.
The editors review papers in the media, profile, and commentary categories. Articles published in the Kentucky Journal of Excellence in College Teaching and Learning reflect the views of the authors, not those of the editors, editorial board, or Eastern Kentucky University.

**Brief Notes on Categories**

**Research**
Educational research is a process of discovery that may or may not use the scientific method. There are limitations in the scientific method and other types of research methodology. The researcher recognizes and defines a problem, formulates a hypothesis, collects data, analyzes the data, and then provides a statement of conclusion that may or may not confirm the hypothesis.

**Theory**
Theory is a tested and testable concept that explains an occurrence. Researchers may analyze existing theories or apply them to contemporary situations. Sometimes researchers develop new interpretations to existing theories. A theory paper involves a lot of critical-thinking, reading, and reflection.

**Practice**
Practice involves almost everything teachers do relative to their profession. Teachers continuously try out new concepts and ideas, and reflect on “best practices.” They usually share their successes with colleagues in the profession.

**Media Review**
We live in an era of information overload and teachers play the role of sorting out and managing the material they intend to use in their classrooms. Media reviews contain reports, demonstrations, and/or critiques of print or electronic resources that enhance or facilitate teaching and learning.

**Profile**
A profile is a feature story of a person that highlights the person’s background, ideas, and accomplishments. Profiles describe qualities that are worthy of emulation.

**Commentary**
A commentary is a piece of writing that gives a perspective on an occurrence or event. It may clarify, explain, or illustrate issues related to the title.
The Kentucky Journal of Excellence in College Teaching and Learning

Mission
The Journal targets the professional development of college/university faculty in the Commonwealth of Kentucky. We also welcome submissions from faculty in the United States and overseas.

Goals
1. To adapt “best practices” to new teaching and learning contexts.
2. To promote innovative research in the scholarship of teaching.
3. To provide a forum for sharing successful teaching/learning strategies.
4. To profile individuals who exhibit excellence in teaching.
5. To stimulate ongoing professional development through teaching/learning resources.

Submissions
Articles must represent innovation in teaching and learning in higher education. Articles must be submitted exclusively to the Kentucky Journal of Excellence in College Teaching and Learning. Manuscripts must be double-spaced with one-inch margins, paragraphs indented five spaces, pages clearly numbered, 12-point Times New Roman font, and should use the most current edition of The Publication Manual of the American Psychological Association. Upload an electronic copy to the Journal website.

Journal Content
The Journal will address innovative and practical teaching resources for faculty. It will also publish informative articles relating to the profession on topics such as:
- “Best practices”
- Innovative environments
- Collaborative practices
- Instructional technology
- Creative pedagogical approaches
- Student retention
- Diversity within the teaching/learning environment
- Creative and innovative teaching tips
- Successful teaching/learning practices
- Profiles of honorary award recipients
- Book and media reviews
- Professional meetings, workshops, conferences and resources

Publication Information
This peer-reviewed electronic Journal is sponsored by the Faculty Development Workgroup/Council on Postsecondary Education and published by the College of Education, Eastern Kentucky University. Reviewers represent faculty from public and private institutions of higher education all over Kentucky and beyond. A print copy of the Journal will be available at the annual Kentucky CPE Faculty Development Conference.
In This Issue

Samuel Hinton • Eastern Kentucky University

Mari Alschuler assessed whether providing faculty in-service training on scoring reflective journals using a rubric would result in statistically significant inter-rater reliability.

Antoinette Davis sought to address the gaps in the research literature concerning online, hybrid, and traditional education.

Adam Akerson examined elementary pre-service teacher candidates’ (PSTCs) perceptions of mathematics, through drawings.

Amanda Joyce studied the relationship between high grades and high teaching evaluations, and their association with other indicators of course difficulty.

Samuel Hinton is Professor of Education, Eastern Kentucky University, and Editor,  
*The Kentucky Journal of Excellence in College Teaching and Learning.*
Facility Inter-Rater Reliability of a Reflective Journaling Rubric  

Mari Alschuler • Youngstown State University

Abstract

There has been a lack of research regarding faculty training in the grading of student reflective journals (RJs). Whether or how one should evaluate RJs remains contentious. This quasi-experimental study assessed whether providing faculty in-service training on scoring RJs using a rubric would result in statistically significant inter-rater reliability. Prior to the study, faculty raters received training on scoring RJs with a rubric based on five levels of reflection. Percent agreement between rater pairs, with 80% set as the inter-rater reliability benchmark, was utilized. Faculty raters scored anonymous BSW and MSW RJs assigned in cultural diversity and oppression courses. Expected learning outcomes included critical and reflective thinking; social justice; application and synthesis of classroom learning to social work practice; ethical awareness; and self-awareness. Fifty percent of RJs collected twice over one term were selected randomly. One faculty pair was selected by chance and assigned under blinded conditions to score either BSW or MSW RJs. Inter-rater reliability of BSW RJ scores ranged from 86% for the first set to 98% for the second set. For the MSW RJs, scores ranged from 85.5% to 83.2%. These findings were all statistically significant and indicated that, with prior training on the purpose of RJs and in using a rubric, faculty may be better able to evaluate RJs fairly.

Keywords: rubrics, social work education, reflective journals, diversity

Clinical educators have their feet in two worlds: professional practice and teaching the next generation of practitioners. Schön (1987) proposed that students in pre-professional programs need to place their learning squarely in the experiential schema, reflecting upon client incidents in order to learn how to function in complex, ever-changing environments.

Students in field placements are expected to apply critical and reflective thinking skills, to develop self-awareness, and begin to work with diverse client populations. These students learn to reflect on action (Freire, 1970/2008; Schön, 1983, 1987) as they begin to assess their underlying beliefs, values, and assumptions about course content, their interactions with clients, and their use of self (Bay & Macfarlane, 2010; Bogo, Regehr, Katz, Logie & Mylopoulos, 2011; Lay & McGuire, 2010; Levine, Kern, & Wright, 2008; McCoy & Kerson, 2013; Urdang, 2010).

It has been reported that students who reflect on a deeper level may be better able to consider their use of self and develop a keener self-awareness (Larrivee, 2008; Marchel, 2004). Urdang (2010) stressed “the importance of incorporating self-reflectiveness into social work education” (p. 525). Learning transferred from the classroom to the practice setting benefits both clinicians and their clients.

Social work educators can benefit from an improved understanding of how to develop, assign, and evaluate reflective writing assignments, and in particular how to create meaningful reflective journal (RJ) assignments that help students integrate course content and apply learning to field practice (Campbell, Schwier, & Kenny, 2009; Cohen, 2010; Taylor & Cheung, 2010). There is a need for critical reflection as students struggle with issues related to discrimination and oppression while they learn about cultural humility and self-awareness.

However, if faculty members have not been trained in reflective practice or RJ, they might not be as well-prepared to foster deeper levels of critical reflection in students (Alschuler, 2012; Dyment & O’Connell, 2010; Hubbs & Brand, 2010). The assignments may feel like busywork or may
not meet their intended goals. Further, faculty may not know how to objectively grade such subjective assignments. Faculty training in reflective practice and journaling is recommended, which led to the development of this study.

**Theoretical Framework**

This faculty development study drew on the theoretical work of Mezirow (1991), Dewey (1933), and Schön (1983, 1987) regarding the roles of transformational learning theory and reflective practice in the context of social work education. Dewey (1933) stated that we learn both from experience and from our reflection on experience—events and the meaning we make of them. Schön (1983) viewed reflection as how one acquires knowledge based on experience.

**Transformational Learning**

Constructionist assumptions about how people create stories about their lives underlie transformational learning theory. We are born into a constructed society with its own set of received meanings. What is transformed is the re-interpretation of past events and behaviors and their accompanying meanings.

Mezirow (1991) posited that transformational learning occurs through critical reflection to address cultural biases and assumptions, misunderstandings, or distortions (Bay & Macfarlane, 2011). In transformational learning theory, the events that occur in people’s lives are less important than how people interpret them (Mezirow, 1991). If a new experience does not fit any prior schema, we may become confused as to how to label, narrate, or categorize it. Through interpretation, we make meaning out of experience (Hoshmand, 2004; Mezirow, 1991).

In clinical education, the process of learning about oneself is central (Hoshmand, 2004). Critical reflection of what is taught in the classroom permits transformation to occur in students. Hoshmand credited critical reflection as one of three elements in transformational counselor education; the other two elements were “critical dialogue and the exercise of critical thinking” (p. 83). Duggan (2005) described the transformational education of adult students as often occurring when a critical incident triggered the identification of differences between the actuality and the ideal.

A sense of disequilibrium may create what Freire (1970/2008) termed conscientization and which Mezirow (1991) called a “disorienting dilemma.” Freire posited that conscientization involves three processes: naming, reflecting, and acting. Plack et al. (2007) described this ‘disorientation’ as a common problem because practitioners and interns regularly “encounter ambiguous, undifferentiated clinical problems that require higher order thinking, not simply recall of knowledge and skills” (p. 286). Meaning-making is involved in transformational learning theory as well as in reflective practice (Fiddler & Marienau, 2008). Through transformational learning practices, including thinking and writing reflectively, students can learn to foster their awareness of the disorienting dilemma as they work to become authentic, reflective practitioners.

**Reflection**

Through reflection, one is able to transform the problem, discover innovative solutions, and develop new skills one might call upon should a future similar circumstance occur (Schön, 1983). Sandars (2009) created a hybrid, transformational definition of reflection which highlights the importance of context:
Reflection is a metacognitive process that occurs before, during and after situations with the purpose of developing greater understanding of both the self and the situation so that future encounters with the situation are informed from previous encounters. (p. 685)

Schön (1983) differentiated between reflection-in-action and reflection-on-action. Reflection-in-action occurs during a situation in which the learner experiences something novel or in states of uncertainty or value conflict. Loughran (2002) differentiated among the different times at which students may write reflectively: anticipatory, retrospective, and contemporaneous. For example, one might use mental rehearsal or role playing with a peer prior to an anticipated event. Because it may be impossible to write in the midst of an event, Loughran highlighted whether there is still time to reflect quickly and change one’s intervention.

Reflection-on-action is a post-hoc review of an event that already occurred (Schön, 1983, 1987). Through a reconsideration of past events, we develop new ways of knowing (Dewey, 1933). Post-hoc journaling might occur through reminiscence, either emotionally or viscerally, by focusing on vivid details or on feelings. Writing after an event also allows one to re-evaluate what happened, what one’s role was, and what one might have done differently (Fiddler & Marienau, 2008).

Reflection can be part of one’s teaching strategy, according to Mann, Gordon, and MacLeod (2009). In a meta-analysis of 29 articles on reflective practice, they summarized that: students benefited from teachers who modeled reflective practice; reflective thinking could be taught or at least encouraged through guided writing prompts and teacher feedback; and that reflection helped students understand both course content and how to integrate new information.

Critical reflection on experiences with different people is one way to increase self-awareness. RJs can also help students develop professional identities as they become acculturated into a new profession (Lay & McGuire, 2010; McGlamery & Harrington, 2007). Student interns are regularly confronted by issues they have never dealt with previously (Fiddler & Marienau, 2008; Sandars, 2009), and they make decisions and use interventions based on what they have previously learned (Mezirow, 1991; Schön, 1983). Clinical educators are appropriately positioned to assist their students in developing these important skills (Balen & White, 2007; Fritschler & Smith, 2009).

**Reflective Journaling**

The effective, intentional use of RJs requires faculty to be familiar with their purpose and how to construct meaningful assignments, and to come to a measured decision about whether and how to evaluate RJ content (Hume, 2009; Larrivee, 2008; Marchel, 2004; O’Connell & Dyment, 2011). Pavlovich (2007) outlined four dimensions of the reflective process as it helps students develop self-awareness: (a) how learning through reflection-in-action occurs within an experience; (b) metacognitive awareness to think about what occurred; (c) mindful awareness about one’s discomfort, uncertainty, or anxiety surrounding the experience, requiring reconsideration of one’s actions and responses; and (d) planned action in response to the experience and one’s reflection through changing one’s behavior or stance.

In one of the few studies of reflective practice among faculty, Larrivee (2008) assessed faculty who reviewed RJs of pre-service teachers. She drew parallels between reflective thinking and conscientization.
(Freire, 1970/2008), as both situate the teacher in a moral and ethical social environment. She maintained that only through self-reflection—questioning one’s own values, and the broader sociopolitical environment—could one become a reflective teacher. Larrivee focused on how teachers can assist students along a four-level reflective thinking continuum, which she based on Mezirow (1991): (a) pre-reflection or non-reflection; (b) surface reflection; (c) pedagogical reflection; and (d) critical reflection.

Levels of Student Reflective Writing

Many studies on the use of RJs have focused on categorizing levels of written reflections, with most including rubrics with three to seven levels (Aukes, Geertsma, Cohen-Schotanus, Zwierstra, & Slaets, 2007; Alschuler, 2012; Bogo et al., 2011; Grossman, 2009; Kember, McKay, Sinclair, & Wong, 2008; Larrivee, 2008; McGlamery & Harrington, 2007; O’Connell & Dyment, 2011; Pavlovich, 2007). Klenowski and Lunt (2008) pointed out that levels of reflection are seen by some as static entities, when they are anything but rigid. They recommended differentiating between “productive and unproductive reflection,” (p. 206) wherein the latter would be superficial and the former would involve higher cognitive skills such as synthesis (Bloom, 1956).

In revising Bloom’s (1956) taxonomy Pintrich (2002) added metacognitive knowledge as a fourth category. Metacognition includes self-reflection and an awareness of one’s own learning style, cognitive strengths and areas for improvement, and how to select certain learning strategies in order to master content and apply theory to practice. Pintrich highlighted the importance of teachers helping “students make accurate assessments of their self-knowledge” (p. 222). As it pertains to reflective practice and self-awareness, metacognitive knowledge has also been posited as related to learning transfer (Pintrich, 2002).

For the purposes of this study, the researcher created a five-level rubric (see Appendix A) for evaluating student RJs to be written at the beginning and end of two diversity courses, one at the BSW and one at the MSW level. The content for the rubric related directly to course content.

The five levels used in this study were: 0=Responding; 1=Reconsidering; 2=Re-evaluating; 3=Reframing; and 4=Reintegrating (Alschuler, 2012). Students who write at the lowest level (Responding) do so superficially; they give the teacher the minimum expected content. Concrete facts are stated, but with no real evidence of reflective or critical thinking.

At the next level, Reconsidering, students are able to step back from events to think about what occurred. They evidence budding awareness that biases and assumptions may have been received from their sociocultural and political milieu. Their writing is somewhat less superficial and displays beginning awareness of self (Alschuler, 2012).

At the Re-evaluating level, students consider the sociocultural and political context in more depth. They display an understanding of how their own and others’ biases, values, beliefs, and assumptions have been received from their environments. Their RJs may show tentative questioning of authority, self-analysis, and inspection of their own beliefs (Alschuler, 2012).

Students writing at the next level, Reframing, explore social justice issues in the context of theory, personal and professional experience, and the sociocultural and political milieu. They are able to consider other points of view. They may openly question authority or the role of their environment in shaping their values and assumptions. There is an awareness of use of
self in professional practice; these students consider how they might act in the future (Alschuler, 2012).

At the deepest level, Reintegrating, students evaluate their received assumptions. They synthesize course material, personal experience, and sociopolitical realities into a developing sense of self. The content displays professional future plans; character or personal growth; and increased self-awareness (Alschuler, 2012).

To grade or not to grade? Identifying, describing, and labeling levels of reflection have been a main concern; however, others have explored the use of questionnaires, templates, or rubrics to evaluate or grade students’ level of reflection (Aukes et al., 2007; Bogo et al., 2011; Grossman 2009; Hume, 2009; Kember et al., 2008; Lay & McGuire, 2010). Yet, a controversy remains: whether or not, and how, to evaluate student RJs (Creme, 2005; Kennison, 2006; Levine et al., 2008; O’Connell & Dyment, 2011; Plack et al., 2007; Sandars, 2009).

Mann et al. (2009) raised the concern that if a teacher does not evaluate RJs, students may not see any value or purpose in taking the time to write them in a thoughtful manner. Dyment and O’Connell (2010) opined that ungraded assignments may be left unwritten, or viewed as unimportant or busywork, and thus completed superficially. Creme (2005) stated that some colleges force faculty to grade all assignments, tying the hands of instructors who may have preferred some latitude in regard to grading RJs; she recommended grading RJs.

Hubbs and Brand (2010) argued about the necessity to grade RJs. They maintained that if these assignments are seen as data—like exams or academic essays—then how the RJ contents will be graded needs to be made explicit. By so doing, instructors may then establish measurable criteria linked to learning outcomes. The authors suggested that a lack of inter-rater reliability may hamper teachers from grading RJs, as there would be concerns about subjectivity. Without effective measures of observable criteria, they argued, assessment and evaluation may be compromised.

Rationale for the Study

Faculty can benefit from learning how to develop, assign, and grade or evaluate RJs that help students integrate content and apply learning to field practice (Campbell et al., 2009; Cohen, 2010; Taylor & Cheung, 2010). However, there has been an overall lack of research regarding faculty training in the use of RJs (Alschuler, 2012; Dyment & O’Connell, 2010; Larrivee, 2008). The present study considered how faculty might score student RJs using a five-level rubric to evaluate inter-rater reliability of the instrument.

The study took place over one semester at one Midwestern state university. Two RJ assignments were created and integrated into the syllabi for a BSW course on Cultural Diversity and an MSW course on Oppression and Cultural Competence. The study focused specifically on inter-rater reliability in the use of a scoring rubric to add to the literature on faculty’s ability to fairly evaluate subjective student RJs on diversity-related themes.

Methodology

Research Design

This quasi-experimental study looked at faculty evaluation of student RJs as they related to course content on oppression, cultural competence, and diverse populations, using a rubric. The research hypothesis was that there would be statistically significant inter-rater reliability among reader/raters’ RJ scores at both the BSW and the MSW level. Fifty percent of RJs were selected randomly twice over one semester, using an Internet-based random
number generator (www.stattrek.com), and scored blindly by faculty who were not instructors for the two courses. Raters were selected using chance (coin flip) to ascertain whether they would read undergraduate or graduate RJs.

**Procedures**

The researcher (PI) received approval from the university’s Institutional Review Board (IRB). Instructors who would be teaching Cultural Diversity and Oppression and Cultural Competence in the fall were informed about the study in advance by the department chair. During this fall term, Cultural Diversity enrolled 27 undergraduates in one section. Oppression and Cultural Competence was held in three sections, for a total of 44 graduate students. The PI emailed the rubric and insert for the syllabi to the instructors.

On the first day that each class section met, the PI personally introduced the purpose of the study to students and informed them that their RJs would be graded by their instructor using the same rubric, and that the external raters’ scores would not be shared with their instructor. The PI emphasized that the outside raters would not know their identity and instructed them on how to create a unique individual identifier, which they were told to place on their RJs.

**Sampling and Recruitment**

The population of faculty raters for this study was drawn from all full- and part-time faculty members teaching in one social work department accredited by the Council on Social Work Education (CSWE) at one Midwestern public university. Faculty members were informed of the study during earlier faculty meetings. The inclusion criterion was their volunteering to participate in an in-service training and agreeing to read and score RJs with the rubric.

Four faculty members agreed to participate: two tenured, full-time professors who had taught undergraduates for an average of 22 years (range 20-24 years), and who had taught graduates for an average of nine years (range 6-12); and two adjunct instructors who had taught undergraduates for an average of 21 years (range 11-31). One part-time instructor had not taught at the graduate level, while the other had taught at the graduate level for nine years.

Under direction of the IRB and due to confidentiality concerns, no other demographic data was collected from raters. The reason for this is that the reader/raters work in a small department and their identities could be revealed should more demographic information be obtained.

The PI paired one full-time and one part-time faculty to read either the BSW or the MSW RJs. A chance method (coin flip) was utilized to select which pair would read which level of RJ; the same pair read the same level both times, and did not know which level they were scoring.

**Faculty In-Service Training**

After signing informed consent forms, raters received a two-hour in-service training by the PI that covered reflective practice, RJs, and the rubric based on five levels of reflection (Alschuler, 2012). They were given sample RJ entries to practice rating using the rubric, and then discussed their perceptions of how they had rated the samples. They requested and received permission to score in-between levels using “.5” (e.g., 2.5, 3.5). Raters were reminded not to discuss their ratings with one another.

**Data Collection**

**Reflective Journal #1**

The first set of RJs (17 of 27 BSW and 37 of 44 MSW RJs) was collected three weeks into the fall term. Some had not been handed
in on time or were missing a student’s identifier code. Students’ names and individual identifiers were entered into a confidential research log, separated by educational level (BSW or MSW). Online random sampling was used to select 50% of the RJs (www.stattrek.com). Nine BSW and 18 MSW RJs were selected for scoring. Blank rubrics were tagged with the students’ identifiers and stapled to the RJs for the raters.

When reading the first set of BSW-level RJs, one of the two BSW raters recognized the writing of one student, did not score it, and returned it to the PI, who also pulled the score for that student from the other rater. Thus, eight, rather than nine, initial BSW RJs were rated. This also had the unintended consequence of making the rater aware that she was grading undergraduate RJs.

Reflective Journal #2

The second set of RJs was collected three weeks before the end of fall term. A total of 38 MSW and 21 BSW RJs were collected for the second set. Using the same procedure, 50% were randomly selected (n=19, MSW; n=10, BSW). RJs were disseminated to the same faculty pairs, who rated them using the attached rubrics.

Findings

Descriptive Information

Time Spent Rating. The faculty members reported that they spent an average of 1.75 hours (range: 1.0 to 2.5 hours) reading and scoring the first set of RJs. The average amount of time spent reading and scoring the second set of RJs was reported to be 1.8 hours (range, 1.2 to 2.5 hours).

Rubric Scoring. Each rubric contained five relevant content areas that students were to include in their RJs: critical and reflective thinking; social justice themes; apply and synthesize classroom learning or theory to social work practice; ethical awareness; and self-awareness. Each of the five items could be scored in a range from 0 to 4 points, for a total score of 20 points. No points were given if the student did not write any content in that area. One point was given for substandard content in each area. Two points were given for adequate content. Three points were given for good content. Four points were given for exemplary content.

Data for all items on a student-by-student, item-by-item, and rater-by-rater was entered into Microsoft Excel. Percent agreement was separately calculated for the BSW-level pairs of ratings and for the MSW-level pairs of ratings. Inter-rater reliability was calculated through the percent agreement method. For the purposes of this study, 60% agreement was considered acceptable and 80% was considered statistically significant (Gay, Mills, & Airasian, 2011). Less than one point apart for each of the five items on the rubric was considered “agreement” for this study.

BSW RJs

First Set. There were eight pairs of ratings for five items, for a total of 40 items. Of those, seven pairs of items were more than one point apart (33 of 40 agreed). Percent agreement between the two raters of the first RJ for the BSW class was calculated at 86%.

Second Set. There were 10 pairs of ratings, for a total of 50 items. Only one pair of items was more than one point apart (49 of 50 agreed). Percent agreement between the raters of the second RJ was calculated at 98%. Both findings were thus statistically significant.

MSW RJs

First Set. Eighteen pairs of ratings were reviewed, for a total of 90 items. Thirteen pairs of items were more than one point apart (77 of 90 agreed). Percent agreement between the two raters of the first RJ for the
three MSW sections was thus calculated at 85.5%.

**Second Set.** Nineteen pairs of ratings were reviewed, for a total of 95 items. Sixteen pairs of items were more than one point apart (76 of 95 agreed). For the second set of RJs, the percent agreement was 83.2%. Both findings were thus statistically significant.

**Discussion**

The faculty raters were in agreement the majority of the time, with more variation among those rating the MSW-level student RJs compared to those reading the BSW RJs. The findings were all statistically significant in terms of inter-rater reliability. At times, reader/raters used the .5 to indicate their hesitance in firmly selecting one category over another. For the most part, the difference between “good” and “exemplary” or between “acceptable” and “good” appeared to be fairly clearly delineated. These findings indicated that, with prior training and practice in using a rubric to grade student journals, faculty may be able to evaluate RJs more fairly.

**Limitations**

Validity may be compromised with all self-reported instruments (Gay et al., 2011). It was expected that faculty members rated RJs independently and without consulting with one another. Because randomization was used to select RJs twice during the term, there was no intention of comparing students’ content from the start to the end of the term, which might be a topic for future study. Faculty self-selected to participate, which may have caused a threat of differential selection. However, the researcher used a coin flip to mitigate any potential bias. The range of time each person reported they spent reading and rating RJs indicated that some spent more time and possibly more effort than others; this may have affected differences in pairs of scores.

**Implications for Social Work Educators**

Social work educators serve not only as professors, but as mentors who have the additional task of preparing students for entering the profession. Reflective practitioners who are also educators may serve as role models and mentors to the students they are socializing into the profession. Training in the helping professions includes clinical internships in the field. The courses selected for this study related to the social work profession’s ethical standards and education goals, including social justice, diversity, and cultural competence. These are suggested as suitable topics for the development of RJs into the curricula.

**Conclusion**

Learning to become a social worker entails acculturation into a profession. McGuire, Lay, and Peters (2009) imparted that clinical educators need to help students learn how to manage complexity, relate theory to practice, and use higher-order cognitive skills in making clinical decisions. Social work practice entails encountering unique, difficult, and unfamiliar situations on a regular basis, clinical social workers need to develop their flexibility, adaptability, and use of self in working with others (Levine et al., 2008). Through RJs, students can learn to reframe their clinical and field experiences to foster self-awareness, empathy, and empowerment (Balen & White, 2007; Fritschler & Smith, 2009). As shown in this study, faculty may learn how to develop, assign, and evaluate RJ assignments through in-service training to help students achieve these learning outcomes.
References


Mari Alschuler is Assistant Professor of Social Work and MSW Program Coordinator, Youngstown State University.
## APPENDIX A
### Reflective Journaling Rubric

<table>
<thead>
<tr>
<th>Level of Reflection:</th>
<th>Reintegrating</th>
<th>Reframing</th>
<th>Re-evaluating</th>
<th>Reconsidering</th>
<th>Responding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>Exemplary 4 pts</td>
<td>Good 3 pts</td>
<td>Acceptable 2 pts</td>
<td>Unacceptable 1 pt</td>
<td>No Credit 0</td>
</tr>
</tbody>
</table>

**Critical & Reflective Thinking**
- Questioning implicit, received assumptions, values & beliefs within socio-cultural-political context
  - Thoroughly evaluated & questioned received or implicit assumptions, values, & beliefs, & compared them to others. Explored aspects of socio-cultural-political context in some detail.
  - Considered values & beliefs in context. Evidence of some questioning &/or comparing own views to those of others. Limited exploration of socio-cultural context.
  - Some awareness of receiving some values from society. No exploration of others’ points of view. No mention of context.
  - Emotional or habitual response. No reflective or critical analysis of own values, those of others, or context.

**Social Justice Themes**
- Oppression, Cultural Diversity & Competence; Underlying Structures of Oppression
  - Full exploration of social justice themes including institutional oppression or racism. Discussed structures underlying oppression.
  - Social justice issues explored in some depth. Mentioned various kinds of oppression, but not broad societal structures.
  - Vague discussion about one social justice issue. Focus on personal identity or membership rather than underlying structures.
  - Broad, vague, superficial mention of social justice issues without regard to context.
  - No exploration or mention of social justice issues.

**Apply & Synthesize**
- Classroom Learning/Theory to Social Work Practice (Field, Volunteer, or Personal Experiences)
  - Synthesis of course material/ theory & field or volunteer work (or personal experience). Identified areas needing more training or experience & described steps to do so (a plan).
  - Experiences were thoughtfully considered in light of classroom learning. Areas needing more training mentioned briefly or with broad, unspecific plans.
  - Personal or field experiences briefly mentioned in light of classroom learning. Need for more training not mentioned or very vague.
  - Personal or field experiences briefly mentioned but not related to classroom learning. No evidence of synthesis. No discussion of learning needs.
  - Student repeated back book learning received information & opinions. No discussion of learning needs.

**Ethical Awareness**
- Struggle with ethical dilemmas/conflicts related to cultural differences; cognitive awareness
  - Described specific struggle with an ethical dilemma. Elicited self-awareness intellectually of coming to terms with oppression.
  - Broad exploration of an ethical dilemma or standards conflict related to cultural differences. Some self-awareness.
  - Minimal mention of ethical standards. Limited self-awareness at cognitive level.
  - Stated they have no dilemmas and no potential conflicts. Lack of self-awareness.
  - No mention of ethics, standards, or struggles. Lack of self-awareness.

**Self-Awareness:** Use of self. Questioning in context of diversity & oppression; personal exploration & growth; affective awareness
- Specific examples of awareness of use of self. Questioning stance. Addressed emotional reaction to materials.
  - Broad awareness of use of self. Discusses how own values & beliefs may be changing, &
  - Some beginning self-awareness. Limited description of feelings as they relate to
  - Minimal or vague/broad personal exploration of use of self. No feeling words used.
  - No personal exploration of use of self. No feeling words used.
<table>
<thead>
<tr>
<th></th>
<th>feelings about that.</th>
<th>discrimination or difference.</th>
</tr>
</thead>
</table>

Total Points Possible = 20  Score: ___  © Alschuler, 2012
Measuring Student Satisfaction in Online Mathematics Courses

Antoinette Davis • Eastern Kentucky University

Abstract

For many years, various colleges and universities have found it difficult to measure student satisfaction in online courses. This study examined the growth of math courses that are delivered in the online format. This study looks to address the gaps in the research literature concerning online, hybrid, and traditional education. In particular, it is the intention of this study to investigate satisfaction and its effect on the performance of students as a result of enrolling in online mathematics courses. Many researchers have sought to find ways to determine student satisfaction in online courses. Satisfaction and performance in distance education have always been seen in comparison with traditional education that implements instruction through face-to-face interactions. This study will extend the comparison to include online and hybrid education. An examination of the research literature shows that researchers measure satisfaction and performance in various ways. This situation may well be responsible for the inconsistencies regarding satisfaction and performance found among empirical studies. Although the present study found that older students were not as satisfied in online mathematics courses as younger students, it is not equipped to investigate the reasons driving their lower satisfaction. Future research should look into possible reasons.

Keywords: student satisfaction, online, mathematics courses

With a growing percentage of university students working part-time or full-time and using technology on a more frequent basis in their daily life, colleges and universities are increasingly supplementing their traditional mathematics courses with online equivalents. Online education using the Internet and information technologies is becoming a popular tool for distance education to better meet students’ needs, interests, learning styles, and work schedules (Lim, Kim, Chen, & Ryder, 2008). However, published studies are not consistent in comparing performance and satisfaction of students in traditional and online instruction (Lim et al., 2008). Various weaknesses in research are responsible for this inconsistency.

This study aimed to improve the quality of educational research on distance education by filling in some gaps (or overcoming some weaknesses) in the research literature. First, this study developed and validated an instrument that measured students’ satisfaction with taking online courses (Tables 1-3). Second, this study explored the relationship among student satisfaction, student performance, and individuals’ characteristics, learning preferences, and online (learning) environment (Tables 4-5). Specifically, this study predicted student satisfaction (measured through the developed instrument) from individuals’ characteristics, learning preferences, and online (learning) environment (Table 6); and predicted student performance from those same variables plus student satisfaction (measured through the developed instrument) (Table 7). This chapter will explain the methods that were used to accomplish these purposes. As a result, this study contributed to a better measurement of student satisfaction in an online environment. This will hopefully help researchers and practitioners better understand the complex relationship among student satisfaction, student performance, and students’ characteristics, learning preferences, and online (learning) environment.

Data Sources

In this study, the target participants were all students who were enrolled in an asynchronous online course, College
Algebra, at a certain Community and Technical College in the Midwest region of the United States ($N = 300$ students). The students in the online course were of mixed age, gender, and ethnicity. Students were invited to participate in the study and did not receive any compensation for participation. Students were surveyed anonymously. Data on students’ characteristics, their learning preferences, and the characteristics of the online (learning) environment were collected in an online survey. Students’ viewpoints on personal feedback, perception of online learning, student-student interaction, student-instructor interaction, and social presence in an online course were also collected (using the developed instrument). The researcher also conducted a pretest and posttest of relevant mathematics knowledge and skills.

**Instrumentation**

The instrument, entitled Satisfaction of Online Learning (SOL), included 24 items embedded in eight components that were developed based on the theoretical framework (see Table 1). The validity of this instrument was established by carefully constructing or selecting items that closely reflect each of the components. The items were developed in this study to isolate certain behaviors that were closely associated with each of the eight factors (components) in Table 1. They were constructed using responses to positive statements. Responsive options for each statement (item) included *Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree* (ranging from 1 to 5 respectively). Students with a higher score indicated more satisfaction to a certain area of a certain factor.

After the construction of the instrument, a pilot was conducted to field-test its functions in the spring of 2013. The instrument was emailed to 15 students in the online course who had one week to work on the instrument. Students were instructed to highlight an option that corresponded most closely to their response to each statement that described a behavior or factor associated with student satisfaction in regards to the online mathematics course. Students were also instructed to answer all items, take notes on anything that caused confusion, and record the time that they needed to complete all items. The result of this pilot served to improve the instrument. The effort helped to answer the first research question: Is it possible to develop a valid and reliable instrument that measures the extent to which students are satisfied with learning mathematics in an online environment?

The formal, comprehensive data collection started in the summer of 2013 with the participation of students in all sections of the asynchronous online course, College Algebra (with consents). At the end of this semester, students were administered (a) SOL, (b) an online survey (measuring individual characteristics, learning preferences, characteristics of online learning environment) (see Appendix B), and (c) a test of mathematics knowledge and skills. To validate SOL, the factorial structure of this instrument was validated through confirmatory factor analysis, and the reliability of this instrument was established by calculating the reliability coefficients of each component and all components as a whole. The online survey used a straightforward design, with questions that collected information about individual characteristics, learning preferences, and online learning environment.

The test of mathematics knowledge and skills covered in the online course (i.e., College Algebra) was given to students within the first two weeks and within the last two weeks of the course so that gains in mathematics knowledge and skills could be measured. The test included multiple choice items and open-ended items, both concerning
mathematics knowledge and skills that students learned in the online course (e.g., operations of addition, subtraction, multiplication, and division). Specifically, various aspects of content included mean price, total price, purchase price, rounding, simplifying, combining like terms, ratio, mixed numeral, length, width, angles, and problem solving. This test had been used for many years in the same course, but as an additional check, an experienced mathematician examined the test for the mathematical correctness of the items and the practical appropriateness of the test for the course (i.e., an expert validation process).

**Measures and Variables**

The online survey had three parts: The first part collected individual data, including gender, age, financial aid (as a measure of socioeconomic status or SES), ethnicity, geographic location, highest mathematics course taken in high school, distance learning experience, working experience, and educational level in college. The second part collected data on students’ learning preferences, including visual learning, aural learning, verbal learning, physical learning, logical learning, social learning, and solitary learning. The third part collected characteristics of the online (learning) environment, including instructional format, what time of day to meet, and the delivery method.

The collected data was used to answer the second and third research questions. The second research question concerned whether there is a relationship between student satisfaction in online mathematics courses and individual characteristics, learning preferences, and characteristics of the online (learning) environment. The third research question concerned whether there was a relationship between students’ performance and satisfaction with regard to learning mathematics in an online environment.

For the second research question, the dependent variable was student satisfaction. The independent variables were individual characteristics of students, their learning preferences, and characteristics of the online (learning) environment. Because it was impossible to randomly select participants in this study (i.e., the sample consisted of volunteers), it was important to include student characteristics in the data analysis.

For the third research question, the dependent variable was student performance in the posttest. The independent variables included student performance in the pretest (functioned actually as a covariate), student satisfaction with online mathematics courses, individual characteristics of students, their learning preferences, and characteristics of the online (learning) environment. The data analysis aimed to compare the importance between student satisfaction, students’ characteristics, their learning preferences, and characteristics of the online (learning) environment with student performance in the online course.

**Statistical Procedures**

The statistical procedure for the validation of SOL closely followed the one used in Shen et al. (2012). It begins with an item analysis to make sure that students were using the full range of the responsive options. This task was performed “by examining the frequencies on the responsive options for each statement” (Shen et al., 2012, p. 9). It proceeded to examine the instrument’s factorial validity. A series of confirmatory factor analyses were performed to examine whether the eight-factor structure identified through the literature review were present within our sample of online mathematics students. Specifically, the eight-factor model was compared with two other models
including the null model and the one-factor model. Comparing the proposed model with the null and one-factor models is a routine procedure in instrument validation (Shen et al., 2012). Model-data-fit statistics included $\chi^2$, SRMR, TLI, CFI, AIC, and BIC (Table 3).

The $\chi^2$ statistic gave an indication of overall fit of the data to the model, with a small $\chi^2$ value indicating a good fit. As one of the absolute measures of fit that does not use an alternative model as the base for comparison, the $\chi^2$ statistic provided only a rough idea about model-data-fit, being quite sensitive to sample size, model size, and variable distribution. The standardized root mean square residual (SRMR) was a much better alternative absolute index. An SRMR value smaller than .08 is considered a good fit (see Hu & Bentler, 1999). The comparative fit index (CFI) and the Tucker-Lewis index (TLI) could be considered as relative measures of fit because they used an alternative model as the basis for comparison. CFI avoids the underestimation of the model-data-fit, often occurring when a sample is small. TLI provides a measure of model-data-fit that is independent of sample size. Because both CFI and TLI measured the proportion of variance explained in relation to the null model, a value greater than .90 indicates a good fit (see Hu & Bentler, 1999).

Lastly, because the models in this study were non-nested ones, information-based estimates were also used to evaluate goodness of fit, including Akaike information criterion (AIC) and Bayesian information criterion (BIC). A best-fitting model had the smallest estimate on both AIC and BIC.

Once the factorial structure was “empirically supported, we combined items within each scale in order to produce the mean and standard deviations for each scale” and this task was “performed by taking the average of valid responses within each scale” (Shen et al., 2012, p. 14-15). Distribution of scale scores were then examined with “two distribution indices: skewness, to make sure that scores were roughly symmetrical around the mean; and kurtosis, to make sure that the distributions were not overly peaked or overly flat” (Shen et al., 2012, p. 15) (Table 4). Finally, Cronbach’s alpha was used as the measure of internal consistency. Reliability analysis was performed on each scale and the instrument as a whole (see Shen et al., 2012) (Table 5). This statistical procedure concluded statistical analysis of the first research question.

For the second research question, a multiple regression analysis was performed with student satisfaction as the dependent variable and variables descriptive of individual characteristics, learning preferences, and online (learning) environment as the independent variables (Table 6). After handling missing data on the dependent variable (i.e., SOL), $N = 102$ students remained for data analysis. For the third research question, a multiple regression analysis was performed, with student posttest performance as the dependent variable and student pretest performance as a measure of prior ability (a covariate by nature) (Table 7). The independent variables were the same as those used in addressing the second research question (i.e., variables descriptive of individuals’ characteristics, learning preferences, and the online learning environment). After handling missing data on the dependent variable (i.e., posttest), $N = 68$ students remained for data analysis.

Because the sample size was relatively small in the case of both research questions, independent variables were first examined individually to test their absolute effects. The absolute effects of a variable refer to the effects of that variable that will occur without the presence of other variables in the statistical model. Variables that are found to have absolute effects are then tested together in the statistical model to see if relative...
effects appear. The relative effects of a variable refer to the effects of that variable that will occur in the presence of other variables in the statistical model. This strategy successfully avoided entering a large number of independent variables together into the regression model (the so-called stepwise approach that is not a sound statistical practice when the regression model runs on a small sample).

**Specification and Validity of SOL Items**

The validity of SOL was established by carefully constructing each item based on empirical evidence or references. That is, empirical evidence or references functioned to provide clues for the wording or description of each item. Each piece of evidence or each reference served as a foundation for the construction of each item in SOL. This approach helped to validate the instrument (SOL) with stronger proof and greater clarity. Table 2 presents the specifications and validations of SOL items in detail.

**Summary of Principal Findings**

The instrument, Satisfaction of Online Learning (SOL), was found to be highly valid and highly reliable. Specifically, both item analysis and scale analysis did not show any abnormal distributional properties of SOL. According to the common comparative practice in confirmatory factor analysis, the eight-factor model represented substantial improvement in model-data-fit over the null model and the one-factor model. Reliability analysis indicated substantially high internal consistency across scales and as a whole instrument.

Multiple regression analysis was performed using students’ satisfaction with online mathematics courses as the dependent variable and variables descriptive of individual characteristics, learning preferences, and the online (learning) environment as the independent variables. All of the independent variables were tested for absolute effects and relative effects. Overall, age demonstrated both absolute effects and relative effects and was considered robustly important to student satisfaction. Younger students were more satisfied with online mathematics courses than older students. Pre-calculus/calculus (vs below pre-calculus) and visual learning showed absolute effects but not relative effects, and were thus considered unimportant to student satisfaction. All other variables did not even show absolute effects on student satisfaction. Therefore, students’ satisfaction was related only to their age.

Multiple regression analysis was also performed with posttest scores as the dependent variable, pretest scores as the covariate, and variables descriptive of individual characteristics, learning preferences, online (learning) environment, and satisfaction with online mathematics courses as the independent variables. None of the independent variables showed absolute effects. Therefore, gains in mathematics knowledge and skills from pretest to posttest in the course were not related to individual characteristics, learning preferences, or the online (learning) environment. Neither were gains related to satisfaction with online mathematics courses.

In sum, SOL as an instrument filled in a significant gap in the research literature for measuring students’ satisfaction with online mathematics courses. It provides a valid and reliable alternative evaluation to traditional course evaluation for colleges and universities to determine student satisfaction in their online courses. Although this study attempted to determine the effects of variables that describe individual characteristics, learning preferences, and the online (learning) environment on student
satisfaction, age was the only significant factor separating student satisfaction. Lastly, this study aimed to examine the relationship between student performance and satisfaction in an online environment. However, students’ gains in mathematics knowledge and skills were not related to their satisfaction (nor to individual characteristics, learning preferences, and the online environment).

Revisiting Research Literature

The present study took the position that IT (information technology) does not bring about a new learning culture independent of pedagogical settings (Blömeke, Muller, & Eichler, 2006; Schulz-Zander, 2005; Tergan, 2003; Vovides, Sanchez-Alonso, Mitropoulou, & Nickmans, 2007). Instead, there is a strong need to describe adequate settings of learning and instruction for all kinds of e-learning (Giest, 2010). The present study attempted to understand the pedagogical settings from three essential aspects (characteristics of individuals, learning preferences, and online environment) that may associate with performance and satisfaction in the online learning of mathematics.

Online Environment

A vehement argument has long been waged, pitting distance education against traditional face-to-face education (Tucker, 2001). There are arguments in the research literature that support the “superiority” of alternative instructional environments. For example, Kendall (2001) asserted that online courses can achieve learning goals and student satisfaction as much as, if not more than, traditional courses. After comparing these three different learning environments, Lim et al. (2008) reported that students in the online learning group and the hybrid learning group have statistically significant higher levels of achievement than students in the traditional learning group and students in the hybrid learning group also have greater satisfaction levels with their overall learning experience than students in the traditional group.

There are also arguments in the research literature that support the “inferiority” of alternative instructional environments. For example, Faux and Black-Hughes (2000) found the largest improvement in performance (from pretest to posttest) for students in the traditional face-to-face environment. Students who prefer traditional environment show a stronger mastery goal orientation and greater willingness to apply effort while learning than students who prefer either online or hybrid environments (Clayton et al., 2010).

The present study did not have separate groups in various online environments; instead, preferences for online learning environments were compared in relation to student performance and satisfaction in the online learning of mathematics. In other words, the present study focused on student preferences for online learning environment (i.e., online vs face-to-face, hybrid vs face-to-face). The results of the present study indicated that students who preferred hybrid instructions were as satisfied with their online learning experiences in mathematics as students who preferred traditional instructions. Meanwhile, students who preferred hybrid instructions gained as much in mathematics knowledge and skills in the course as students who preferred traditional instructions. These conclusions hold true to the comparisons between online instructions and traditional instructions. That is, students who preferred online instructions were as satisfied with their online learning experiences in mathematics as students who preferred traditional instructions, and students who preferred online instructions gained as much in mathematics knowledge
and skills in the course as students who preferred traditional instructions. Based on the above findings, this study could not support either the superiority or inferiority of both hybrid instructions and online instructions over traditional instructions from the perspectives of student performance and satisfaction in the online environment of learning mathematics. In particular, the pretest and posttest design of the present study added important insights into the research literature because comparisons based on the longitudinal perspective have been rather rare in the research literature.

**Individual Characteristics**

The limited research literature on individual differences in online learning focuses mainly on age and gender differences. Previous research indicated significant gender differences in performance, attitudes, motivation, and experiences (Ashby, Sadera, & McNary, 2011; Branden & Lambert, 1999; Chen, 1999; Muilenberg & Berge, 2005; Owens, 1998). Previous research also found age to be a significant factor for learning (educational) outcomes in online courses (Ashby et al., 2011; Muilenberg & Berge, 2005; Rekkedal, 1983).

In the present study, age was found to be robustly important to satisfaction with online mathematics courses but unimportant to performance in online mathematics courses. Furthermore, gender differences were not found in either performance or satisfaction concerning the online learning of mathematics. These findings all represent new contributions to the field of online mathematics education. In particular, Thurmond, Wambach, Connors, and Frey (2002) asserted that student satisfaction is influenced by instructional decisions and actions in the online environment but not by student characteristics. The present study suggests that certain individual characteristics (e.g., age) may still have influence on student satisfaction.

**Learning Preferences**

The research literature on online education contains some information on what learning preferences (styles) fit better to the online learning environment such as active vs reflective, sensing vs intuitive, visual vs verbal, and sequential vs global (Kim & Moore, 2005). Schellens and Valcke (2000) noticed that developers of online courses tend to favor visual, applied, spatial, social, and creative styles of learning. Nevertheless, how learning preferences relate to performance and satisfaction remains an under-research issue, which partially motivated the present study.

There are conflicting results regarding whether learning preferences (styles) relate to academic performance (Fahy & Ally, 2005). Some studies on online learning suggest that students’ learning preferences are associated with their course performance (Douzenis, 1999; Sabry & Baldwin, 2003; Terrell, 2002). Meyer (2003) argued that visual learners are more academically successful than aural and kinesthetic learners in an online learning environment (see also Ozbas, 2008 for gender differences in academic performance in an online learning environment that emphasizes visual learning). On the other hand, Santo (2001, 2006) found no relationship of learning preferences to both course grades and test scores.

According to Henry (2008), the visual-verbal dimension of students’ learning preferences (styles) correlates positively with satisfaction as learners in a hybrid (e-blended) course delivery mode but negatively with satisfaction as learners in a traditional course delivery mode. Overall, however, Kearsley (2000) indicated no relationship between students’ learning preferences and their satisfaction with online courses.
The present study provided some further insights into the relationship of learning preferences to performance and satisfaction in the online learning environment. Specifically, learning preferences were related to performance and satisfaction in the online learning of mathematics. Confidence is high in the present study in that satisfaction was measured with a validated instrument and performance was measured in a pretest and posttest design. These features of the present study are rather uncommon in the research literature, permitting the present study to make unique contributions to the current understanding of the relationship between performance and satisfaction.

**Implications**

**Instrument Application**

Kane, Williams, and Cappuccini (2008) argued that student institutional satisfaction surveys are a valuable source of data for instructional improvement but little has been used outside their immediate management improvement purposes. Meanwhile, researchers have commonly used a single-item rating scale to assess student satisfaction, but this approach fails to recognize the complexity of students’ reactions to educational service (Elliott & Shin, 2002). The instrument (SOL) that has been validated in the present study can help improve both situations in that SOL is a great tool to generate specific information on many aspects of student institutional satisfaction that can be easily applied to instruction as well as management of online courses. All of the eight scales within the instrument can be used either individually or collectively to measure student satisfaction for various purposes of instruction and management.

**Age Factor**

The present study found that older students tended to be less satisfied with online mathematics courses than younger students. This finding may serve as a call for instructors to be more attentive to the way that they communicate information to older students in an online classroom. Moore (1993) suggested that for distance learning to be successful, instructors need to pay attention to three elements of transactional
distance theory (dialogue, structure, and learner autonomy) in order to reduce the “distance” experienced by students. When distance is felt by students in the online course, they tend to feel isolated and may stop participating in the subsequent learning activities. The best way to reduce distance is to structure the course in such a way that all learners (both young and old) can benefit from the material that is presented in the online mathematics course. According to Chao and Davis (2001), there are many facets to the online success of math courses such as paying attention to the design and utilization of effective online pedagogy, maintaining active communication between students and the instructor, encouraging interaction between students in the classroom, and using computer programs like Excel as a way to illustrate statistical concepts in the classroom.

In addition, it is important to identify characteristics of students who feel successful with their online learning experiences so as to provide necessary information for instructors and admission officers to either encourage or discourage a student from registering for an online course (Wojciechowski & Palmer, 2005). The present study, in this sense, is useful to administrations at colleges and universities. Younger students are more likely to be satisfied with taking mathematics courses in the online environment than older students can become a factor to aid decision making.

**Limitations**

Sampling-related issues represent the major limitations of the present study. The initial sample size of 259 students was promising, but the three separate data collection procedures (SOL; online survey of individual characteristics, learning preferences, and online environment; mathematics test in pretest and posttest format) produced missing data. As a result, the confirmatory factor analysis was based on 123 students with valid SOL scores. Confirmatory factor analysis based on such a sample size is less ideal (see Shen et al., 2012). Missing data reduced sample size again when it came to answering the second and third research questions. Multiple regression analysis to address the second research question was based on 102 students, and that to address the third research question was based on 68 students. Although the strategy of examining absolute effects individually first is effective and sufficient analytically, results regarding the second and third research questions need to be considered tentative. Due to the limited number of online students that can often be reached in any study, it is suggested that future researchers accumulate data from different semesters to improve the number of student responses (Kuo, 2010).

The use of volunteer sample represents another major limitation. Although the difficulty in obtaining a random sample is adequately realized in educational research, a large number of studies based on volunteer samples need to be conducted for any meaningful synthesis of results across studies. It is suggested that future researchers continue this line of research with various volunteer samples if random sampling is impractical. Indeed, several researchers have suggested that more research be done to collectively deal with the lack of large random samples concerning online learning (e.g., Ertmer et al., 2007; Kuo, 2010; Richardson, 2005).

The scope of the present study was limited. The part of the online survey that collected information on individual characteristics was not as comprehensive as one would like. For example, Dabbagh (2007) found that intrinsically motivated learners with a positive attitude toward the instructor and a high expectation for grades and degree completion are more likely to
succeed in a distance education course. The space limitation prevented the present study to look into whether students’ attitude and expectation can predict performance and satisfaction in the online learning of mathematics. This issue leaves sufficient opportunities for future researchers.

Recommendations for Further Research

Although some recommendations for further research have been offered in the previous section, more discussion on this line of research may be beneficial. SOL is a valid and reliable instrument, but nevertheless it was developed based on a particular college-level mathematics course (i.e., College Algebra). Therefore, this instrument needs to be validated and even modified within and beyond the area of mathematics education. For example, SOL can be validated for more advanced mathematics courses taught in an online environment; and SOL can also be validated for college science courses. Although it is reasonable based on the review of research literature to expect SOL to be a general measure of satisfaction with any online courses, further validation is necessary.

Because of the tentative nature of the results from multiple regression analyses, there is a need for future researchers to replicate studies concerning the comprehensive relationship among student performance and satisfaction in online learning of mathematics as well as individual characteristics, learning preferences, and online (learning) environment. Following a similar logic, further studies may include different variables descriptive of individual characteristics, learning preferences, and online (learning) environment.

Although the present study found that older students were not as satisfied in online mathematics courses as younger students, it is not equipped to investigate the reasons why they are less satisfied. Future research can look into possible reasons. Some research may even focus on older students and their reasons for taking math courses online. As a result, future online courses can be built with more resources and help so that their time in the online environment may become a good experience.

References

Teacher Education International Conference 2001 (pp. 1348-1349). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).


Antoinette Davis is Lecturer, Department of Mathematics and Statistics, Eastern Kentucky University.
Table 1

*Foundation for Instrument Development*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness of feedback</td>
<td>1-3</td>
</tr>
<tr>
<td>Timeliness of feedback</td>
<td>4-6</td>
</tr>
<tr>
<td>Use of discussion boards in the classroom</td>
<td>7-9</td>
</tr>
<tr>
<td>Dialogue between instructors and students</td>
<td>10-12</td>
</tr>
<tr>
<td>Perception of online experiences</td>
<td>13-15</td>
</tr>
<tr>
<td>Instructor characteristics</td>
<td>16-18</td>
</tr>
<tr>
<td>Feeling of a learning community</td>
<td>19-21</td>
</tr>
<tr>
<td>Computer-mediated communication</td>
<td>22-24</td>
</tr>
</tbody>
</table>

Table 2

*Distribution of Responses and Descriptive Statistics across Items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>.10</td>
<td>.05</td>
<td>.19</td>
<td>.35</td>
<td>.32</td>
<td>3.74</td>
<td>1.23</td>
</tr>
<tr>
<td>Q2</td>
<td>.11</td>
<td>.04</td>
<td>.18</td>
<td>.33</td>
<td>.34</td>
<td>3.76</td>
<td>1.26</td>
</tr>
<tr>
<td>Q3</td>
<td>.10</td>
<td>.05</td>
<td>.20</td>
<td>.34</td>
<td>.31</td>
<td>3.72</td>
<td>1.23</td>
</tr>
<tr>
<td>Q4</td>
<td>.11</td>
<td>.02</td>
<td>.19</td>
<td>.34</td>
<td>.35</td>
<td>3.81</td>
<td>1.23</td>
</tr>
<tr>
<td>Q5</td>
<td>.12</td>
<td>.02</td>
<td>.17</td>
<td>.29</td>
<td>.40</td>
<td>3.83</td>
<td>1.31</td>
</tr>
<tr>
<td>Q6</td>
<td>.09</td>
<td>.05</td>
<td>.22</td>
<td>.34</td>
<td>.30</td>
<td>3.70</td>
<td>1.20</td>
</tr>
<tr>
<td>Q7</td>
<td>.08</td>
<td>.12</td>
<td>.37</td>
<td>.21</td>
<td>.21</td>
<td>3.35</td>
<td>1.18</td>
</tr>
<tr>
<td>Q8</td>
<td>.09</td>
<td>.08</td>
<td>.28</td>
<td>.28</td>
<td>.27</td>
<td>3.55</td>
<td>1.22</td>
</tr>
<tr>
<td>Q9</td>
<td>.09</td>
<td>.09</td>
<td>.22</td>
<td>.33</td>
<td>.27</td>
<td>3.60</td>
<td>1.22</td>
</tr>
<tr>
<td>Q10</td>
<td>.10</td>
<td>.03</td>
<td>.23</td>
<td>.28</td>
<td>.37</td>
<td>3.78</td>
<td>1.25</td>
</tr>
<tr>
<td>Q11</td>
<td>.10</td>
<td>.06</td>
<td>.24</td>
<td>.32</td>
<td>.29</td>
<td>3.65</td>
<td>1.23</td>
</tr>
<tr>
<td>Q12</td>
<td>.10</td>
<td>.05</td>
<td>.30</td>
<td>.28</td>
<td>.28</td>
<td>3.59</td>
<td>1.22</td>
</tr>
<tr>
<td>Q13</td>
<td>.08</td>
<td>.07</td>
<td>.25</td>
<td>.32</td>
<td>.28</td>
<td>3.66</td>
<td>1.22</td>
</tr>
<tr>
<td>Q14</td>
<td>.09</td>
<td>.07</td>
<td>.19</td>
<td>.37</td>
<td>.28</td>
<td>3.68</td>
<td>1.21</td>
</tr>
<tr>
<td>Q15</td>
<td>.13</td>
<td>.08</td>
<td>.27</td>
<td>.19</td>
<td>.33</td>
<td>3.51</td>
<td>1.37</td>
</tr>
<tr>
<td>Q16</td>
<td>.12</td>
<td>.09</td>
<td>.26</td>
<td>.24</td>
<td>.28</td>
<td>3.48</td>
<td>1.32</td>
</tr>
<tr>
<td>Q17</td>
<td>.12</td>
<td>.09</td>
<td>.27</td>
<td>.25</td>
<td>.26</td>
<td>3.44</td>
<td>1.30</td>
</tr>
</tbody>
</table>
Table 3
Results of Model Data Fit from Confirmatory Factor Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>X²</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null factor</td>
<td>5052.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 factor</td>
<td>1474.15</td>
<td>0.74</td>
<td>0.72</td>
<td>0.06</td>
<td>6277.32</td>
<td>6479.80</td>
</tr>
<tr>
<td>8 factor</td>
<td>590.71</td>
<td>0.92</td>
<td>0.90</td>
<td>0.05</td>
<td>5449.88</td>
<td>5731.10</td>
</tr>
</tbody>
</table>

Note. Values other than means and SDs represent percentages.

Table 4
Descriptive Statistics across Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness of Feedback</td>
<td>3.74</td>
<td>1.24</td>
<td>-0.92</td>
<td>0.01</td>
</tr>
<tr>
<td>Timeliness of Feedback</td>
<td>3.78</td>
<td>1.25</td>
<td>-0.99</td>
<td>0.15</td>
</tr>
<tr>
<td>Use of Discussion Boards</td>
<td>3.50</td>
<td>1.21</td>
<td>-0.54</td>
<td>-0.49</td>
</tr>
<tr>
<td>Dialogue between instructors and students</td>
<td>3.67</td>
<td>1.23</td>
<td>-0.77</td>
<td>-0.19</td>
</tr>
<tr>
<td>Perceptions of online experiences</td>
<td>3.62</td>
<td>1.26</td>
<td>-0.69</td>
<td>-0.40</td>
</tr>
<tr>
<td>Instructor characteristics</td>
<td>3.44</td>
<td>1.31</td>
<td>-0.44</td>
<td>-0.81</td>
</tr>
<tr>
<td>Feeling of a learning community</td>
<td>3.29</td>
<td>1.30</td>
<td>-0.31</td>
<td>-0.85</td>
</tr>
<tr>
<td>Computer-mediated communication</td>
<td>3.22</td>
<td>1.27</td>
<td>-0.20</td>
<td>-0.83</td>
</tr>
</tbody>
</table>

Table 5
Reliability Statistics across Scales

<table>
<thead>
<tr>
<th>Scales</th>
<th>Number of Items</th>
<th>Reliabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness of Feedback</td>
<td>3</td>
<td>0.98</td>
</tr>
</tbody>
</table>

34
Timeliness of Feedback 3 0.98
Use of Discussion Boards 3 0.98
Dialogue between instructors and students 3 0.98
Perceptions of online experiences 3 0.98
Instructor characteristics 3 0.98
Feeling of a learning community 3 0.98
Computer-mediated communication 3 0.98
Instrument as a whole 24 0.98

Table 6 
Multiple Regression Results Estimating Effects of Individual Characteristics, Learning Preferences, and Online Environment on Satisfactory with Online Mathematics Courses

<table>
<thead>
<tr>
<th>Variables</th>
<th>Absolute Effect</th>
<th>SE</th>
<th>Relative Effect</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (continuous)</td>
<td>-.87*</td>
<td>.24</td>
<td>-.59*</td>
<td>.28</td>
</tr>
<tr>
<td>Male (vs female)</td>
<td>-2.9</td>
<td>6.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (vs non-White)</td>
<td>3.20</td>
<td>8.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-calculus/calculus (vs below pre-calculus)</td>
<td>16.37*</td>
<td>8.11</td>
<td>15.84</td>
<td>8.80</td>
</tr>
<tr>
<td>Up to associate degree (vs high school diploma)</td>
<td>.40</td>
<td>5.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor and beyond (vs high school diploma)</td>
<td>-13.19</td>
<td>6.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial aid (vs no financial aid)</td>
<td>-7.26</td>
<td>7.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of working experience (continuous)</td>
<td>-.55</td>
<td>.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of online courses (continuous)</td>
<td>.24</td>
<td>2.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning preferences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual learning (continuous)</td>
<td>4.98*</td>
<td>2.41</td>
<td>3.51</td>
<td>2.65</td>
</tr>
<tr>
<td>Aural learning (continuous)</td>
<td>-1.26</td>
<td>2.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal learning (continuous)</td>
<td>3.31</td>
<td>2.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical learning (continuous)</td>
<td>2.32</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logical learning (continuous)</td>
<td>2.89</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social learning (continuous)</td>
<td>3.94</td>
<td>2.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solitary learning (continuous)</td>
<td>-2.93</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Online environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference on online (vs face-to-face)</td>
<td>9.67</td>
<td>6.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7

* p < .05.

Multiple Regression Results Estimating Effects of Individual Characteristics, Learning Preferences, Online Environment, and Satisfactory with Online Mathematics Courses on Gains in Mathematics Performance

<table>
<thead>
<tr>
<th>Variables</th>
<th>Absolute Effect</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (continuous)</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Male (vs female)</td>
<td>.71</td>
<td>.65</td>
</tr>
<tr>
<td>White (vs non-White)</td>
<td>1.29</td>
<td>.86</td>
</tr>
<tr>
<td>Pre-calculus/calculus (vs below pre-calculus)</td>
<td>.163</td>
<td>.75</td>
</tr>
<tr>
<td>Up to associate degree (vs high school diploma)</td>
<td>-.12</td>
<td>.61</td>
</tr>
<tr>
<td>Bachelor and beyond (vs high school diploma)</td>
<td>.55</td>
<td>.65</td>
</tr>
<tr>
<td>Financial aid (vs no financial aid)</td>
<td>1.30</td>
<td>.86</td>
</tr>
<tr>
<td>Years of working experience (continuous)</td>
<td>.01</td>
<td>.08</td>
</tr>
<tr>
<td>Number of online courses (continuous)</td>
<td>.39</td>
<td>.26</td>
</tr>
<tr>
<td><strong>Learning preferences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual learning (continuous)</td>
<td>-.20</td>
<td>.23</td>
</tr>
<tr>
<td>Aural learning (continuous)</td>
<td>-.01</td>
<td>.33</td>
</tr>
<tr>
<td>Verbal learning (continuous)</td>
<td>-.31</td>
<td>.30</td>
</tr>
<tr>
<td>Physical learning (continuous)</td>
<td>-.07</td>
<td>.25</td>
</tr>
<tr>
<td>Logical learning (continuous)</td>
<td>-.03</td>
<td>.26</td>
</tr>
<tr>
<td>Social learning (continuous)</td>
<td>-.10</td>
<td>.29</td>
</tr>
<tr>
<td>Solitary learning (continuous)</td>
<td>-.19</td>
<td>.30</td>
</tr>
<tr>
<td><strong>Online environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference on online (vs face-to-face)</td>
<td>.19</td>
<td>.80</td>
</tr>
<tr>
<td>Preference on hybrid (vs face-to-face)</td>
<td>-.115</td>
<td>1.18</td>
</tr>
<tr>
<td>Scheduled sessions (vs non-scheduled sessions)</td>
<td>-.55</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Asynchronous (vs synchronous)</td>
<td>-.43</td>
<td>.61</td>
</tr>
<tr>
<td>Satisfactory with Online Mathematics Courses</td>
<td>-.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

* $p < .05$
Preservice Teachers’ Perceptions of Mathematics through Drawings

Adam Akerson • Stephen F. Austin State University

Abstract
Research indicates that mathematics anxiety is particularly high in pre-service teachers (Bekdemir, 2010; Gresham, 2007; Hembree, 1990). These future teachers will soon be entering classrooms of their own, responsible for teaching mathematics to young children, who need strong teachers. A 2013 report from The National Assessment of Educational Progress (NAEP), found that 42% fourth-graders performed at or above the proficient level in math (NCES, 2013). The purpose of this study was to examine elementary pre-service teacher candidates (PSTCs) perceptions of mathematics, through drawings. Drawings were analyzed before and after a semester-long field experience in a constructivist mathematics environment. The participants included 56 PSTCs with a field placement in K-5 settings. The researcher used open-coding to evaluate the pre- and post-field experience drawings. Three overarching themes emerged from the data: (1) emotions related to mathematics, (2) the mathematics environment, and (3) experiences within the mathematics environment. Over the course of the field experience semester, the number of negative and isolated perceptions of mathematics decreased. The results of the study hold implications for teacher preparation programs in planning field experience placements for their teacher candidates. Purposeful field placements can allow PSTCs to evaluate and inform their own understanding of mathematics, and learn pedagogical strategies to benefit their future students.

Keywords: elementary, pre-service teacher, perceptions, mathematics, drawings

Everyone seems to come from a different mathematical background. Some are enthused with the prospect of putting problem solving into action, while others harken back to an isolated, anxiety filled classroom. How do you visualize mathematics? This study sought to gain insight into elementary pre-service teacher candidates’ perceptions of mathematics through drawings. The pre-service teacher candidates (PSTCs) in this study will soon be entering classrooms of their own, responsible for teaching mathematics to young children. In 2013, a report from The National Assessment of Educational Progress (NAEP) found that 42% of fourth-graders performed at or above the proficient level in math (NCES, 2013). One possible means of increasing the performance of mathematics learners is by preparing teachers who are comfortable with their own mathematical abilities and capable of delivering instruction. Teachers must know and understand deeply the mathematics they are teaching and be able to use that knowledge in their teaching (NCTM, 2014). It is crucial for those preparing future teachers to understand the perceptions that exist towards mathematics, particularly when sensitivities may be present.

Literature Review

Mathematics Experiences
People’s understanding of mathematics, along with their ability to solve problems, are shaped by the teaching they were afforded throughout their education (NCTM, 2014). Pre-service teacher candidates bring with them a variety of experiences related to mathematics. Most of these occurrences come from time spent as a K-12 student. Unfortunately, some PSTCs may have experienced anxiety related to mathematics from their times as a student. Research reveals that mathematics anxiety is particularly high in pre-service teachers (Bekdemir, 2010; Gresham, 2007; Hembree, 1990). Furthermore, those who have negative experiences in a mathematics classroom tend to have higher anxiety related to mathematics, and these experiences tend to
increase as one advances through school (Bekdemir, 2010). For those choosing to become educators, previous mathematics experiences may influence their preparation as teachers.

Baloğlu and Koçak (2006) categorize mathematics anxiety into three major groups: situational, dispositional, and environmental. Situational causes of anxiety relate to the circumstances in which mathematics is being used. Dispositional anxiety includes personality-related factors one brings to the setting. Lastly, environmental causes of mathematics anxiety include prior perceptions, attitudes and experiences that have impacted an individual. Some PSTC may have experienced one or more of these anxieties, which makes analyzing the perceptions that exist an important component of teacher educator programs.

A teacher can also have an impact on their students’ anxiety level toward mathematics (Vinson, 2001). Educators’ anxiety towards mathematics may not only impact their students, but their ability to teach math and their confidence in how much they know. Anxiety can also influence the decisions teachers make about what to teach and how often, particularly in early grades (Geist, 2015). Fortunately, pre-service teacher candidates’ attitudes towards mathematics can be changed, even over a short period of time (Hodges & Kim, 2013).

**Mathematics Instruction & Learning**

The learning environment has the ability to strongly influence the outcomes of students (LaRocque, 2008). The learning environment is not just influential to children, but also pre-service teachers. Protheroe (2007) suggests that an effective math environment includes teachers who:

- Project a positive attitude toward mathematics and about students’ ability to “do” mathematics.
- Demonstrate acceptance of students’ divergent ideas.
- Influence learning by posing challenging and interesting questions.

Teachers have the ability to change attitudes by using effective instructional techniques, focusing on what students are able to do, encouraging multiple outcomes, and being sensitive to previous frustrations and failures (Furner & Duffy, 2002).

Pre-service teachers should be provided with a variety of examples related to concepts in mathematics, which can allow for individual growth (Bekdehmir, 2010). Teacher candidates are better able to understand mathematics concepts and procedures when they are presented on the pictorial or concrete level. What they understand, they are able to teach (Vinson, 2001). Additionally, research from Gresham (2007) suggests that pre-service teachers prefer doing mental math or working with others to solve problems.

The process of thinking about math mentally encourages learners to build on number relationships to solve problems rather than relying on memorized procedures (Parrish, 2010). To many of us, it may be hard to imagine mathematics in a classroom setting where the use of paper and pencil, calculators, and textbooks are not the focal point of instruction. Yet, people do not often have access to such tools when encountering, say, a 25%-off sale at a department store. Indeed, most individuals perform these calculations mentally. Unfortunately, in an age of assessment and accountability, learners are often taught to “show” their work according to teacher-prescribed procedures. However, mental math encourages learners to concentrate on number relationships and use these relationships to develop efficient, flexible strategies with accuracy, rather than hold quantities and procedures in their heads (Parrish, 2010).
Mathematics Discourse

Mathematical discourse involves conversations about mathematics. Discourse provides a means of communicating ideas and clarifying understanding. Through communication, ideas are reflected upon, refined, discussed, and amended (NCTM, 2000). The quality of dialogue between teachers and learners, and among learners, is of crucial importance to learning and educational attainment (Mercer & Sams, 2006). Furthermore, when students are challenged to communicate their thinking to others, they must be clear, convincing, and precise in their use of mathematical language (NCTM, 2000).

A teacher’s role in establishing a classroom of discourse is not to be a disseminator of information. Rather, the teacher is responsible for establishing an environment in which discourse can take place. The National Council for Teachers of Mathematics (2000) suggests that a teacher of mathematics can facilitate discourse by:

- Posing questions and tasks that elicit, engage, and challenge students’ thinking.
- Listening carefully to students’ ideas.
- Asking students to clarify and justify their ideas.

In addition, creating an environment that accepts, respects and considers all student responses encourages discussion (Parrish, 2010). Listening to others’ explanations provides learners with opportunities to develop their own understanding (NCTM, 2000). When discussing a problem, supplying a variety of solutions may create mathematical conflict, allowing students to consider the context of a problem and the reasonableness of the solution (Cengiz, 2013). This allows students to engage in the process of mathematics. Student explanations should include mathematical rationales and arguments, not just descriptions of procedures (NCTM, 2000). When students are engaged in mathematical dialogue of rationale, they do not easily back down from their thinking, and often must be convinced by their peers, which requires discourse to take place. Conversations that allow for mathematical ideas to be explored from different perspectives help participants refine their thinking and make connections (NCTM, 2000). Noticeably absent from this type of mathematical engagement is a list of teacher-prescribed procedures for solving problems.

Perceptions through Drawings

In education, drawings have been commonly used to evaluate children’s feelings or perceptions. These drawings can be especially important in regulating emotions (Drake & Winner, 2013). Additionally, drawings are considered a technique that can provide insight into a subject’s inner world, rather than relying on verbally expressed perceptions (Regev & Ronen, 2012). In elementary settings, drawings are often used as a tool to help foster writing. Research suggests that drawings are an effective form of rehearsal for narrative writing, and in some cases more beneficial than discussion (Caldwell & Moore, 1991).

Vygotsky (1978) suggests that drawing is a pictorial language, allowing children to find concrete visual means of representing their thoughts, much like its own form of speech. These visual means of representing thoughts hold implications for mathematics. When a student gains access to mathematical representations, and the ideas they express, they can create images that capture mathematical concepts or relationships. This process allows students to acquire a set of tools that can significantly expand their capacity to model and interpret physical, social, and mathematical phenomena (NCTM, 2014).

The Campaign for Drawing suggests that drawing is a vehicle for helping both children
and adults understand the world. Drawings can also be used to help think, feel, shape, and communicate ideas (www.campaignfordrawing.org). Drawings have previously been used as data in exploring perceptions among adults, including those in the field of healthcare (Mays et al., 2011). Research on drawings with adults in education, including pre-service teachers, has been much more limited. The few instances of such research have often been connected to a teacher preparation methods course (Burton, 2012; Lee & Zeppelin, 2014; Rule & Harrell, 2006) rather than field experience. Rule and Harrell (2006) found that the use of drawings could help students connect to previously unconscious images of mathematics, and in so doing, shifts the mathematics anxiety complex toward a more positive framework. In a similar study, conducted with pre-service teacher candidates enrolled in a mathematics methods course, Burton (2012) suggests having PSTCs draw math at the end of an internship experience to see how perceptions of mathematics have changed.

**Methods**

The participants for this study included 56 pre-service teacher candidates enrolled in a field experience course for elementary education, all of which were female. The participants were working toward an EC-6 teaching endorsement from the state in which they resided. Participants came from five different sections of a field experience course. As part of the course, participants were placed in a K-5 university charter school classroom four days a week, three hours each day, over the course of twelve weeks. The field experience included a one-hour lab that met once a week to discuss a variety of components related to their placements. Furthermore, PSTCs were enrolled in a science, social studies, and online mathematics methods course in concurrence with their field experience placement.

As part of the field experience requirements, PSTCs were required to teach lessons related to math, science and social studies. All grade-level placements included number talks associated with mental math as part of their daily mathematics. Parrish (2011) describes number talks as 5-15 minute conversations around specific computation problems solved mentally, also referred to as mental math. During these talks, students are able to communicate their thinking and justify solutions to problems performed mentally. The goal of number talks is to develop more accurate, efficient, and flexible problem-solving strategies. PSTCs observed number talks daily and were responsible for leading whole group number talks as a required teaching component of the field experience.

**Data Collection**

Prior to PSTCs being placed in their field experience setting, they were required to attend three days of orientation. The orientation included information about the objectives of the field experience, discussion of the course timeline, and the kind of professionalism expected in the classroom. It was during orientation, prior to placement in the field, that the pre-field experience drawings were collected. Participants were informed verbally and in writing that their participation in the study was voluntary. PSTCs were specifically asked, “What do you think of when you hear the word mathematics?” Participants were then instructed to draw their impressions of mathematics. PSTCs were told that their impressions might include previous experiences, emotions, or any other image visualized when the term mathematics was heard. After being provided an opportunity to draw impressions of mathematics, PSTCs
were asked to write a few sentences to describe or clarify their drawing. Participants were asked not to filter images. The researcher assured participants that their drawings were confidential. These same procedures were followed in the collection of post-field experience drawings, at the completion of the field experience semester.

**Data Analysis**

The drawings and descriptions of the pre- and post-field experience drawings were open-coded by the researcher in an attempt to closely examine any similarities and differences (Mertens, 2005) between drawings. There were three overarching themes that emerged from the data: (1) emotions related to mathematics, (2) the mathematics environment, and (3) experiences within the mathematics environment.

In the first category, the researcher coded drawings based on emotions related to mathematics. Emotions were categorized as positive, negative, or neutral. Specific drawings, like smiling or frowning faces, were determining factors in classifying emotions. A drawing was categorized as neutral when neither a positive or negative emotion was specified. Participants’ written descriptions regarding particular emotions were also used to categorize drawings.

The second category of coded images relates to the environment in which mathematics was represented. These images depicted mathematics in a classroom, real world, or more abstract setting. Classroom images contained visuals like problems on a whiteboard, books, tests, etc. Real-world images were represented by images in which math was portrayed outside of the classroom, like finding the distance of two objects. People are able to use their knowledge flexibly when they have an understanding of mathematics. However, students who memorize facts or procedures without a mathematical understanding may have difficulty applying their knowledge (NCTM, 2000). Abstract environments were those that did not reference a specific environment. Examples of abstract images include symbols of mathematical operations (addition, subtraction, multiplication, division) or manipulatives associated with mathematics, like math cubes.

Upon further examination of the mathematics environment, another category became evident: the experiences within the mathematics environment. Based on these experiences, the drawings and descriptions were coded as either “collaborative” or “isolated” mathematics. Collaborative images included children working with one another. These images may have also included a description in which words like “sharing” or “discussion” was evident. Isolated images referred to those in which one individual was depicted, or multiple individuals were present but not interacting, such as sitting in rows of desks. A neutral image was coded as one that did not reference a collaborative or isolated mathematics experience, such as the picture of an equation or mathematical symbol.

**Results**

**Pre-Field Experience Drawings**

PSTCs’ initial drawings and writings indicated a variety of experiences related to mathematics. A negative experience or image was reflected in 46% of PSTCs’ initial drawings. These negative images took a variety of forms. Many of the participants included a student with a frown on their face (see Figure 1) or a question mark above their head. In addition to visual representations of negative images, student descriptions further revealed a negative association with mathematics. One PSTC stated, “I work really hard to understand math concepts and
it just doesn’t click with me. I get frustrated.” Another participant stated, “To me math is very difficult. I always feel behind or like I am the only person who does not understand. Math feels like a foreign language.” Other words that PSTC used consistently to describe their images included: “stressful,” “nervous,” “confusing,” and “frustrating.”

Ten PSTCs expressed a positive experience or image related to mathematics. Many of these images include a smiley face or an illuminated light bulb. In the coding of descriptions, the researcher identified a few common words (e.g., enjoy and understand) along with phrases, such as “I can do it!” One teacher candidate stated, “I have always enjoyed math. I enjoy challenging myself and thinking about things from different angles” (See Figure 2). Positive images and descriptions represented 18% of participants’ drawings.

Sixteen PSTCs (27%) composed drawings that included mathematical symbols or operations (see Figure 3). Examples of these images included base ten blocks, number lines, or mathematical operations. Participant descriptions of these drawings referenced terms like “graphing”, “counters”, “multiplication”, and “division.” Since these images and descriptions did not indicate a positive or negative experience with mathematics, they were categorized as “neutral” drawings.

To further examine the drawings, the researcher categorized the drawings based on the setting depicted in the drawings and descriptions. Of the initial drawings, twenty-two PSTCs (39%) provided images or descriptions that represented mathematics in a classroom setting. These images included visuals like a white board, children in rows of desks, textbooks, or a teacher. Forty-six percent of participants’ images and descriptions were abstract representations of math, such as mathematic symbols or operations, while eight PSTCs made connections to mathematics in the real world, like in a pet store (see Figure 4).

Further investigation of the mathematics environment led the researcher to categorize experiences within the environment as isolated, collaborative, or neutral. Nine percent of PSTCs’ drawings or descriptions were coded as collaborative. These images included pictures of children in a circle, or talking through the use of word bubbles (see Figure 5). Collaborative descriptions included words like “sharing” or “discussion.” Forty participants (71%) drew images or provided descriptions that suggested an isolated mathematics-learning environment. Isolated drawings may have included children sitting in a desk by themselves (see Figure 6). A neutral drawing was indicative of 20% of participants. A neutral image did not indicate a collaborative or isolated experience, such as a picture of a mathematical symbol.

Post-Field Experience Drawings

PSTCs’ final drawings and descriptions also indicated a variety of experiences related to mathematics. A positive experience or image was reflected in 36% of PSTCs’ post-field experience drawings (see Table 1). Examples of these images included children and/or teachers smiling, some with light bulbs above their heads. One PSTC included a continuum of six faces progressing from unhappy to happy. The descriptions included with the drawings also indicated positive experiences related to mathematics. One PSTC stated, “My outlook on math is completely different from the beginning of the semester.” Another PSTC stated, “I drew myself smiling instead of frowning because math does not scare me anymore.” Other words that PSTCs used frequently to describe their images include: “happiness”, “understanding”, and “passion”. Over the course of the study, the number of images indicative of a positive emotion grew by
18%, while the number of negative emotions decreased by 39% (See Table 1). One participant included the following description: “Before the field experience, I hated math with a passion. I was not good at it. After this experience and being immersed in this new math, I have a better understanding of math. I am more confident in my abilities.”

To further examine the drawings, the researcher categorized the drawings based on the setting depicted in the drawings. These images included classroom, real world, and more abstract settings. From the pre-field experience drawings, to the post-field experience drawings, the number of images associated with the classroom environment increased by 13% (See Table 2), while the number of images reflecting real-world settings decreased. The number of drawings coded as an abstract setting remained unchanged. One participant described the setting in this way: “Math is not just done in rows of desks or with paper and pencil. It can be done mentally by playing games, using books, during lunch, and through other activities in-class and outside of class.”

Drawings were also coded by type of mathematics interaction: collaborative, isolated, or neutral. The number of images coded as collaborative increased by 39%, while drawings with an isolated experience saw a decline of 43% (See Table 3). Illustrations in which neither a collaborative or isolated experience was depicted saw a decline of 11%. One PSTC stated, “In our class, math is very interactive and openly communicated. Students’ minds are constantly moving as they work different strategies and find different solutions.” Another PSTC stated, “Learning is a social process. Children can learn from peers and teachers, through discussions and seminars.”

**Analysis**

Pre-service teacher candidates will soon be teaching in classrooms of their own, which makes it vitally important for these future teachers to recognize and understand their own attitudes towards mathematics. The results from the drawings and descriptions associated with emotions related to mathematics, collected in this study, are consistent with other research related to pre-service teachers’ attitudes and perceptions (Bekdemir, 2010; Gresham, 2007; Hembree, 1990). Pre-field experience data revealed that 46% of participants held a negative perception of mathematics. Further analysis revealed that these negative perceptions were often accompanied by mathematics experiences in which the participant may have felt isolated, or alone, without support to figure out a math problem.

Post-field experience drawings indicate a negative emotion associated with four of the 56 participants in the study. In comparing pre- and post-field experience data, drawings representative of a negative emotion saw a decrease of 39%. Further analysis revealed that the number of drawings indicative of an isolated mathematics experiences, decreased by 31%. Drawings that depicted a classroom setting increased by 13%, while a decrease of the same percentage was found in the number of drawings illustrating mathematics in the real world.

The coding of the descriptions accompanying the mathematics drawings revealed that 15 participants (27%) referenced the term “mental math”. None of the descriptions that referenced mental math accompanied a drawing associated with a negative emotion. The feeling towards mental math may have been influenced by the field experience placement. PSTCs observed number talks centered around mental math on a daily basis and were required to facilitate a number talks lesson in their classroom placement.
Discussion

Based on the results of this study, the field experience setting seemed to influence PSTCs’ perceptions of mathematics. In particular, allowing PSTCs to observe and facilitate instruction related to mental computation, or mental math, through number talks seemed to leave an impression. The majority of PSTCs left this experience with a positive emotion connected to mathematics, which was not the case prior to the field experience. PSTCs observed classroom settings in which students were encouraged to ask questions of each other, rather than look to the teacher for the answer. Consequently, a number of PSTCs’ post field-experience drawings reflected an environment in which students were free to share their experiences without judgment or fear of providing an incorrect response. Overall, the perceptions towards mathematics seemed to shift from a negative isolated experience toward a positive and collaborative environment.

The results of this study provide implications for educators in teacher preparation programs, especially those working with PSTCs in field experience settings. Providing field experiences, which allow PSTCs to observe mathematics in a setting where collaboration and discussion are critical components, like a number talks, may prove beneficial in aiding future teachers’ understanding of mathematics. Additionally, quality field experiences may help PSTCs better understand how to establish collaborative learning opportunities, free from math anxiety, for their future students.

The results of this study also hold implications for current mathematics educators. There may be countless teachers currently teaching who themselves have a negative perception of mathematics. These teachers may be providing the same anxiety-filled experiences they were afforded as children. Examining the perceptions and attitudes of educators in the field may help administrators identify teachers who could benefit from continued training and professional development opportunities involving mathematics.

In follow-up activities related to the PSTCs’ perception of mathematics in field placements, it would be beneficial to investigate participants’ content knowledge related to mathematics. Shulman (1986) indicates that subject matter used to be of great importance in the preparation of teachers. This was replaced with an emphasis on pedagogy, culture, and policies. The purpose of this study was not to identify content knowledge, although a few participants alluded to math content. For example, one participant stated “Math use to frustrate me and now it makes more sense to me.” From this statement, it is not clear what exactly makes “more sense”, the math content itself, or methods for teaching mathematics. Follow-up interviews with participants could help clarify such information.

Although not all future educators will have positive perceptions of mathematics, there is hope. Quality field experience placements, in which mathematics is an emphasis, can aide PSTCs in gaining valuable teaching experience. These same field placements may also help build confidence along with a better understanding of concepts and instruction related to mathematics.
References


Table 1
*Emotions Related to Mathematics*

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Pre-drawings</th>
<th>Post-drawings</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>10 (18%)</td>
<td>20 (36%)</td>
<td>10 (18%)</td>
</tr>
<tr>
<td>Negative</td>
<td>26 (46%)</td>
<td>4 (7%)</td>
<td>-22 (39%)</td>
</tr>
<tr>
<td>Neutral</td>
<td>20 (36%)</td>
<td>32 (57%)</td>
<td>12 (21%)</td>
</tr>
</tbody>
</table>

*Note.* A neutral image was representative of a drawing in which neither a positive nor a negative emotion was indicated.

Table 2
*Mathematics Setting*

<table>
<thead>
<tr>
<th>Representation</th>
<th>Pre-drawings</th>
<th>Post-drawings</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom</td>
<td>22 (39%)</td>
<td>29 (52%)</td>
<td>7 (13%)</td>
</tr>
<tr>
<td>Real World</td>
<td>8 (14%)</td>
<td>1 (2%)</td>
<td>-7 (13%)</td>
</tr>
<tr>
<td>Abstract</td>
<td>26 (46%)</td>
<td>26 (46%)</td>
<td>No Change</td>
</tr>
</tbody>
</table>

*Note.* An abstract image was representative of a drawing in which neither a classroom nor a real-world image was indicated, such as a mathematics operation symbol.

Table 3
*Mathematics Experience*

<table>
<thead>
<tr>
<th>Interactions</th>
<th>Pre-drawings</th>
<th>Post-drawings</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative</td>
<td>5 (9%)</td>
<td>27 (48%)</td>
<td>22 (39%)</td>
</tr>
<tr>
<td>Isolated</td>
<td>40 (71%)</td>
<td>24 (43%)</td>
<td>-16 (32%)</td>
</tr>
<tr>
<td>Neutral</td>
<td>11 (20%)</td>
<td>5 (9%)</td>
<td>-6 (11)</td>
</tr>
</tbody>
</table>

*Note.* A neutral image was representative of a drawing in which neither a collaborative nor an isolated experience was indicated.
Figure 1. PSTC-generated drawing depicting a negative impression of mathematics. Characteristics distinguishing this drawing as a negative drawing include the frowning face, low grade, and thunderstorm of mathematics.
Figure 2. PSTC generated drawing depicting a positive impression of mathematics. Characteristics distinguishing this drawing as a positive image associated with mathematics include the smiling face and acknowledgment of enjoyment.
Figure 3. PSTC-generated drawing depicting a neutral impression of mathematics. Neither a positive or negative experience is evident from the drawing. Characteristics in this neutral image associated with mathematics include the use of shapes and mathematical operations.
Figure 4. PSTC-generated, real-world image associated with mathematics. Characteristics in this real-world image associated with mathematics include the reference to locations beyond the classroom such as a grocery store, law office, and candy store.
Figure 5. PSTC-generated drawing depicting the collaboration of mathematics. Characteristics depicting collaboration in this drawing include the two students working together while problem solving.
Figure 6. PSTC-generated drawing depicting the isolation of students in mathematics. Characteristics depicting isolation in this drawing include the students sitting in rows unengaged with each other and appearing to be direct taught.

Adam Akerson is Assistant Professor at Stephen F. Austin State University.
Course Difficulty and its Association with Student Perceptions of Teaching and Learning  RESEARCH

Amanda Joyce  •  Murray State University

Abstract
Grade inflation has long been an issue in academia, and with this comes the concern that instructors will feel pressured to inflate grades in order to improve student evaluations of their teaching. Many historical studies have demonstrated associations between higher grades and higher teaching evaluations. The purpose of this investigation was to determine the relationship between high grades and high teaching evaluations, and their association with other indicators of course difficulty. Anonymous, end-of-semester, teaching evaluations were collected from 156 students in 6 sections of 3 unique courses in the Psychology department of a large Southeastern University between 2011 and 2014. Students were asked to report on various aspects of their learning experience, including their instructor’s effectiveness, the level of mutual respect in the classroom, and their expected grade in the course, among other variables. Students’ agreement with the statement, “Overall, the instructor’s teaching was effective,” positively related to their evaluation of all individual aspects of the instructor’s effectiveness (e.g., “The instructor was well-prepared”; “The instructor presented subject matter clearly”; all $r$’s > .433; all $p$’s < .001). However, student evaluations of overall instructor effectiveness showed no association with their expected grade in the course ($r = .133$, $p = .101$), nor with the number of writing assignments or exams given by the instructor (all $r$’s < .138; all $p$’s > .088). The results imply that instructors need not feel pressured to reduce course demands in order to improve student evaluations.

Keywords: grade inflation, course difficulty, instructor evaluations, student satisfaction

Grade inflation has long been of concern in academia, and with this comes the concern that instructors will feel pressure to inflate grades and to otherwise create easier courses in order to improve student evaluations of teaching. The purpose of this investigation was to determine if the association between high grades and high teaching evaluations exists today, as it has in the past, while also examining if individual indicators of course difficulty relate to students’ perceptions of courses.

Concerns about grade inflation can be traced back to the 1970’s (Bowers, 1970; Juola, 1976). Grade inflation occurs when student grades improve, but student achievement does not (Stone, 1995). Though students are happy to receive high grades in their courses, it is problematic to educators as well as to their future employers, because it compresses all grades at the top of a spectrum such that it is difficult to tell the best students from those who are only good, and the good students from those who are only mediocre (Johnson, 2006). Unfortunately, there is ample evidence of grade inflation during the late 20th century. One study investigating the impact of grade inflation from 1962 to 1985 found an increase in average grade point average from 2.49 to 2.93 (Sabot & Wakeman-Linn, 1991). Similarly, the percentage of students expecting an A or A- grade in a course increased by 10% across the 1990’s (Eiszler, 2002).

These increasing grades are encouraged by a student populace who rewards universities and instructors for an artificially inflated grade point average. During the 1990’s, as the number of students who expected an A or A- in a course increased, the average rating on student teaching evaluations also increased by .1 points (Eiszler, 2002). Indeed, other research supports the notion that students give more favorable course evaluations to instructors of easier courses, and that they preferentially enroll in sections of courses that are known to be easy (Cohen, 1981; Feldman, 1989; Johnson, 2006).
To instructors who need favorable evaluations, such as those fighting for tenure and those hoping to renew one-year contracts, the temptation to create easier classes for the purpose of more favorable student evaluations can be difficult to resist. In fact, studies show that faculty are known for trying to influence student evaluation scores (Simpson & Siguaw, 2000). This is corroborated by evidence that adjunct faculty – temporary members of the faculty who run the risk of being replaced quickly – give higher grades to students than do more permanent faculty members (Sonner, 2000). Faculty are also more prone to providing students with higher grades than average in an environment where a “student-as-customer” viewpoint is more strongly endorsed (Stone, 1995).

Many would argue, however, that artificially increasing grades in this way is a disservice to students. Though learner-centered, active learning, the type often found in more challenging courses, can be uncomfortable to students, most agree that it is beneficial to their learning (Weimer, 2002). For example, discussion within courses is associated with better attainment of higher-order knowledge (Garside, 1996), and the amount of time spent studying outside of the classroom relates to academic achievement (McFadden & Dart, 1992). Do students, especially those who are viewed as customers, not deserve a classroom environment that provides them with a more thorough education?

For this reason, the purpose of this investigation is to examine if we can correct this classroom anomaly by providing evidence that artificial grade inflation may not necessarily impact students’ evaluation of courses and instructors. Much of the literature surrounding grade inflation and student course evaluations is rooted in the late 20th century. Therefore, the current study provides important information about whether previous associations still exist today, nearly a decade or more after some of the most influential studies were published.

Previous research suggesting that easy courses produce stronger student evaluations frequently measured course ease in terms of students’ grades. Some research has focused on other measures of difficulty, such as course workload, but they generally did so in terms of students’ perceptions of this workload, rather than objective measures, such as the number of assignments or exams given per semester (Gillmore & Lowell, 1994; Greenwald & Gillmore, 1997; Marsh & Roache, 2000). Therefore, the current investigation adds to our understanding of the topic by not only updating the field on current trends regarding the association between grades and student evaluation scores, but also examining a wider range of variables regarding course difficulty than has previously been explored.

Method

Anonymous, end-of-semester teaching evaluations were collected from 156 students in six sections of three unique courses in the Psychology department of a large Southeastern University between 2011 and 2014. Students were asked to report on various aspects of their learning experience, including their instructor’s effectiveness, their ability to think independently about course material, the level of mutual respect in the classroom, and their expected grade in the course, among other variables. Data were also collected regarding the number of exams and assignments given in each course.
Procedure

End of Semester Evaluations
Students completed end-of-semester evaluations for each class in which they were enrolled. These evaluations were innate to the course and were issued university-wide for all courses. Students received emails prompting them to complete the evaluations online. If students did not complete these evaluations before a university-chosen deadline, they received daily emails reminding them to complete the evaluations. Students also received emails from their instructor on the first and eighth days of each evaluation period. Said emails encouraged students to provide their honest feedback on the course so as to help the instructor improve the course for students in future semesters. The evaluation period each semester closed two weeks after students received the initial university email announcing the evaluation’s availability. Fifty-eight percent of students, across the 6 sections, completed their evaluations.

For this evaluation, students were asked to rate their level of agreement with a number of statements on a Likert-type scale (1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree). They also provided open-ended feedback on the course and, finally, reported their expected grade in the course on a GPA-like scale (4 = A, 3=B, 2=C, 1=D, 0=F).

Of interest to this investigation are Likert-type variables relating to instructor effectiveness and students’ experience of the course itself. As such, even though students reported on variables relating to the physical environment of the classroom, such as “How would you rate the physical environment in which you took this class, based upon your ability to see, hear, concentrate, and participate?,” these variables were not included in the analyses. The full text of the variables included in the analyses can be found in Table 1 below, under the subheading “Evaluation Questions”.

Course Difficulty Variables
Data were also collected regarding the number of exams, as well as the number of in-class and out-of-class assignments, administered each semester. Exams consisted of 40-50 multiple choice questions each, and each exam was administered during a single class-period, with no other class discussion happening on exam days. In-class assignments were defined as those assignments that were assigned, completed, and submitted within a single class period, with no expectation of out-of-class effort to occur on the assignment. In-class assignments could be completed in groups and were discussed at a class-wide level upon submission, during the same class period. Out-of-class assignments, or homework, were those assignments that were completed outside of the classroom and submitted on the course website associated with each class.

Results

Descriptive Statistics
Descriptive statistics for all variables of interest in this study can be found in Table 1. The reported maximum on each evaluation question score matched the maximum possible score of each scale, while the reported minimum only sometimes matched the minimum possible score of the scale. Means also fell toward the top of the scales. Numbers of in-class activities ranged between a low of 1 during one semester to a high of 21 during a different semester, while number of out-of-class homework assignments remained consistent at 2, and tests varied between 3 and 4 administrations per semester.
Pearson Correlations

Evaluation Responses and Difficulty of the Course. To test the hypothesis that the difficulty of a course would correspond to changes in student responses to instructor evaluations, Pearson correlations were calculated between evaluation questions and a student’s expected grade, and the numbers of in-class activities and exams administered throughout the semester. The expected grade in the course was included as a “course difficulty” variable for the purpose of this analysis, in order to examine if students’ perception of grade impacted their assessment of the course. The number of out-of-class homework assignments was excluded from this analysis, because it did not vary from semester to semester. Table 2 presents the results of the analysis.

Course difficulty variables rarely related to course evaluation responses. The expected grade in the course positively related to responses on only three evaluation questions (all other r’s < .154, all other p’s > .057), and the number of in-class activities assigned positively related to only a student’s reported ability to think independently about course material (all other r’s < .128, all other p’s > .114). In contrast, the number of tests assigned in a course negatively related to five separate variables (all other r’s < -.151, all other p’s > .061). Expected grade, in-class assignments, and number of exams administered were all unrelated to students’ agreement with the statement “Overall, the instructor’s teaching was effective” (all r’s < .138; all p’s > .088).

Evaluation Responses and Overall Course Effectiveness. In order to determine if the above lack of correlations was indicative of a true lack of association or, rather, a lack of cohesiveness in student responses, Pearson correlations were also calculated between students’ ratings on individual items on the evaluation and students’ rating of overall teaching effectiveness. This teaching effectiveness variable is a single item on the evaluation (rather than a calculated average of multiple variables), but was chosen for this analysis because it is frequently used as a single number meant to represent an instructor’s effectiveness at the institution at which these evaluations were given. As can be seen in Table 2, this effectiveness rating related to responses on all other evaluation questions (all other r’s > .433, all other p’s < .001).

Discussion

Over the last half-century, various reports have described how grades in college courses have risen despite students reporting less time spent studying (e.g., de Vise, 2012). One potential explanation for this trend is that instructors face pressure to maintain strong student evaluation scores and believe that they can achieve higher scores by decreasing the difficulty of their course and artificially inflating grades. Historical research has found that this tactic may be well founded—that there is an association between assigning high grades and earning high student evaluation scores. However, this research has limited application to today’s academic climate, as some of the most important studies were conducted more than a decade ago. Moreover, very few focus on variables of course difficulty that can be objectively reported on by the instructor.

Thus, the purpose of this investigation was to determine if a course’s perceived difficulty level still shows an association with student evaluation scores today. In the current investigation, each measurement of course difficulty showed some association with students’ ratings of overall instructor effectiveness. Meanwhile, the strength of associations between these difficulty measures was not of the same magnitude as were the associations between
individual items of the evaluation. As such, the evidence suggests that instructors need not feel pressure, as they did in the past, to reduce course demands in order to improve student evaluations.

**Course Difficulty and Teaching Evaluations**

One’s expected grade in a course, as well as the number of assignments and exams given in each course, did impact ratings on some individual items of the course evaluation, though these correlations were relatively small and sporadic. Students’ expected course grades, for example, did not relate to any items related to their view of the instructor. Instead, the expected grades related to items about themselves and their understanding of the material, such as “I have improved my ability to think independently about course material.” Perhaps students are praising their instructor indirectly through these responses, but it is also possible that students who earn higher course grades actually do have a stronger ability to think independently about course material than do students who do more poorly in the class, and that the association between grades and evaluation items is warranted here and not indicative of grade inflation.

Similarly, students who completed more in-class assignments also reported a better ability to think independently about course material. If easier courses were still a strong predictor of high teaching evaluations, one would not expect this finding. Expecting more active work from students in this way can cause some grumblings (Weimer, 2002), yet some characteristic of these in-class assignments actually increased students’ reports of independent thinking. One possible explanation of this could be the increased class discussion that resulted from each of these assignments, yet previous research suggests that students give lower evaluation scores in courses that involve more active learning such as this (Lake, 2001). Perhaps, then, the positive association between number of assignments and student ratings can be explained by the feedback that accompanied these assignments, as submissions were returned to students with written comments on their work. Students appreciate feedback on assignments, indicating that they are motivated by more than just a grade, and those who receive personalized feedback in a course are more satisfied than those who do not (Gallien & Oomen-Early, 2008; Higgins, Hartley, & Skelton, 2002).

The negative association between number of exam scores and various teaching evaluation items is the only one that would have been predicted by previous research relating easier classes to higher evaluations. When students were given more exams, they reported less positive characteristics of their own learning as well as less positive characteristics of the instructor and class environment. One possible explanation for these negative associations is that exams were administered for full class periods, meaning that those courses with more exams had fewer class periods to discuss course material. More exams may also cause more test anxiety, which is known to lead to lower course performance (Hill & Wigfield, 1984; Maehr & Midgley, 1991) and perhaps lower evaluations. Perhaps the negative association lies in the nature of the exams themselves. All exams in this investigation were multiple choice with no written feedback given by instructors. Students believe that multiple choice tests measure a lower form of knowledge and adjust their study techniques accordingly (Scouller, 1998). Thus, when they are given more multiple choice exams throughout the semester, they may feel less need to deeply engage in the course and learn material, which explains the negative association between exams given and course evaluations.
With all of this being said, it is important to note that the strength of associations between course difficulty items and evaluation items is low. What’s more, none of the measures of course difficulty in this investigation related to the item, “Overall, the instructor’s teaching was effective”, which is the item most frequently used from this array to concisely describe an instructor’s ability. Therefore, this investigation reveals that in today’s society, it may not be beneficial to instructors to artificially decrease the difficulty of their course in an effort to receive higher course evaluation scores. With the pressure to artificially inflate grades removed, perhaps instructors can begin to better serve their students with more challenging courses. This transition away from the current trend will be difficult for many, as universities and courses will not have uniform levels of course difficulty, and as students in more challenging courses will have lower grades than their grade-inflated peers, thus making them less competitive for the job market (Johnson, 2006; Sabot & Wakeman-Linn, 1991). Still, instructors need to consider the benefits of creating more academically rigorous courses for their students. Over the course of the last 50 years, at the same time as students’ grades steadily rose, the amount of time spent studying steadily fell, and the number of students making no gains in critical thinking throughout college rose (Arum & Roksa, 2011; de Vise, 2012). By increasing the rigor of courses, perhaps we can counteract these negative consequences of grade inflation.

Limitations and Future Directions
This investigation provides important evidence that instructors need not feel pressured to artificially decrease the difficulty of their courses in order to improve student evaluation scores. However, some limitations need to be considered. First, the correlation of individual evaluation items was quite high. Taking this in tandem with the fact that scores on most items skewed toward the top end of the distribution, there may be a halo effect in students’ evaluations. Second, this investigation, relative to more prolific historical investigations, involved a relatively small sample size across a relatively homogenous sample of courses and students. More research is necessary to determine if these same effects can be found in a wider span of classes and students in which students may provide more negative course evaluations. Third, though this investigation provided evidence that a larger number of multiple choice exams in a course may relate to more negative ratings on some evaluation items, more research is necessary to determine if this association holds true across other varieties of exams (short answer, essay, mixed, etc.). Finally, because scores on evaluation items in this investigation were relatively high, future research should determine if some environments are more likely than others to discourage the historical association between high grades and high teaching evaluations. For example, perhaps students are more forgiving of rigorous academic environments when instructors demonstrate high support for students and create an environment of mutual respect.

In the meantime, the results imply that instructors under pressure of limited contracts need not artificially decrease the difficulty of their courses, but should instead work to give their students a better learning environment. Classes can be made more challenging as long as students can be provided proper feedback on assignments, and as long as exams are not given at the expense of more thorough course discussion. Perhaps in the future, instructors across disciplines and universities can find a solution to the damage that grade inflation has caused. We need to work together to help students learn effectively while remaining
competitive for graduate school and the job market.

References
Table 1
Descriptive Statistics for Key Study Variables

<table>
<thead>
<tr>
<th>Evaluation Questions</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructor was well prepared.</td>
<td>3</td>
<td>6</td>
<td>5.77</td>
<td>.546</td>
</tr>
<tr>
<td>The instructor presented the subject matter clearly.</td>
<td>2</td>
<td>6</td>
<td>5.74</td>
<td>.645</td>
</tr>
<tr>
<td>The instructor provided feedback intended to improve my course performance.</td>
<td>2</td>
<td>6</td>
<td>5.66</td>
<td>.727</td>
</tr>
<tr>
<td>The instructor fostered an atmosphere of mutual respect.</td>
<td>4</td>
<td>6</td>
<td>5.84</td>
<td>.398</td>
</tr>
<tr>
<td>I have a deeper understanding of the subject material as a result of this course.</td>
<td>3</td>
<td>6</td>
<td>5.62</td>
<td>.733</td>
</tr>
<tr>
<td>My interest in the subject matter was stimulated by this course.</td>
<td>1</td>
<td>6</td>
<td>5.40</td>
<td>1.099</td>
</tr>
<tr>
<td>Overall, the instructor’s teaching was effective.</td>
<td>2</td>
<td>6</td>
<td>5.62</td>
<td>.724</td>
</tr>
<tr>
<td>I improved my ability to think independently about the course material.</td>
<td>2</td>
<td>6</td>
<td>5.34</td>
<td>.725</td>
</tr>
<tr>
<td>I learned to identify problems and explore different solutions.</td>
<td>2</td>
<td>6</td>
<td>5.29</td>
<td>.856</td>
</tr>
<tr>
<td>The instructor used a scholarly approach in presenting content</td>
<td>2</td>
<td>6</td>
<td>5.58</td>
<td>.703</td>
</tr>
<tr>
<td>The instructor treated students with respect.</td>
<td>4</td>
<td>6</td>
<td>5.86</td>
<td>.397</td>
</tr>
<tr>
<td>The instructor was effective in administering the class and organizing materials.</td>
<td>2</td>
<td>6</td>
<td>5.74</td>
<td>.615</td>
</tr>
<tr>
<td>The grade I expect in this course is.</td>
<td>2</td>
<td>4</td>
<td>3.54</td>
<td>.550</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course Difficulty Variables</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of In Class Activities</td>
<td>1</td>
<td>21</td>
<td>7.97</td>
<td>8.114</td>
</tr>
<tr>
<td>Number of Homework Assignments</td>
<td>2</td>
<td>2</td>
<td>2.00</td>
<td>.000</td>
</tr>
<tr>
<td>Number of Tests</td>
<td>3</td>
<td>4</td>
<td>3.73</td>
<td>.445</td>
</tr>
</tbody>
</table>

Notes. N’s range from 153 to 156 due to occasional missing data. Evaluation questions were rated on a 6-point Likert-type scale, though responses on some scales did not show this same range of variability. Expected course grade is student-reported and follows a typical 4.0 grading scheme (4 = A, 3 = B, …0 = F).
Table 2

Pearson Correlation of Evaluation Items with Indicators of Course Difficulty

<table>
<thead>
<tr>
<th>Evaluation Questions</th>
<th>Effective</th>
<th>Grade</th>
<th>ICA</th>
<th>Exams</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructor was well prepared.</td>
<td>.818***</td>
<td>.137</td>
<td>.086</td>
<td>-.129</td>
</tr>
<tr>
<td>The instructor presented the subject matter clearly.</td>
<td>.769***</td>
<td>.119</td>
<td>.128</td>
<td>-.179*</td>
</tr>
<tr>
<td>The instructor provided feedback intended to improve my course performance.</td>
<td>.697***</td>
<td>.154</td>
<td>.118</td>
<td>-.150</td>
</tr>
<tr>
<td>The instructor fostered an atmosphere of mutual respect.</td>
<td>.657***</td>
<td>.086</td>
<td>.124</td>
<td>-.167*</td>
</tr>
<tr>
<td>I have a deeper understanding of the subject material as a result of this course.</td>
<td>.840***</td>
<td>.180*</td>
<td>.081</td>
<td>-.136</td>
</tr>
<tr>
<td>My interest in the subject matter was stimulated by this course.</td>
<td>.805***</td>
<td>.204*</td>
<td>.091</td>
<td>-.151</td>
</tr>
<tr>
<td>Overall, the instructor’s teaching was effective.</td>
<td>--</td>
<td>.133</td>
<td>.095</td>
<td>-.138</td>
</tr>
<tr>
<td>I improved my ability to think independently about the course material.</td>
<td>.599***</td>
<td>.177*</td>
<td>.171*</td>
<td>-.223**</td>
</tr>
<tr>
<td>I learned to identify problems and explore different solutions.</td>
<td>.577***</td>
<td>.159*</td>
<td>.123</td>
<td>-.177*</td>
</tr>
<tr>
<td>The instructor used a scholarly approach in presenting content.</td>
<td>.699***</td>
<td>.041</td>
<td>.155</td>
<td>-.202*</td>
</tr>
<tr>
<td>The instructor treated students with respect.</td>
<td>.433***</td>
<td>.038</td>
<td>.027</td>
<td>-.059</td>
</tr>
<tr>
<td>The instructor was effective in administering the class and organizing materials.</td>
<td>.601***</td>
<td>.148</td>
<td>.036</td>
<td>-.071</td>
</tr>
</tbody>
</table>

Notes. N’s range from 153 to 156 due to occasional missing data. “Effective” = item related to overall rating of instructor effectiveness. “Grade” = Expected course grade is student-reported and follows a typical 4.0 grading scheme (4 = A, 3 = B, …0 = F). “ICA” = number of in-class assignments. “Exams” = number of tests administered. Number of out of class homework assignments was not included in this analysis, as the number did not vary from semester to semester. * = p< .05, ** = p < .01, *** = p< .001.

Amanda Joyce is Assistant Professor, Murray State University.