Development and Validation of an Assessment Instrument for Course Experience in a General Education Integrated Science Course

Juhong Christie Liu,[1,a] Kristen St. John,[2] and Anna M. Bishop Courtier[3]

ABSTRACT

Identifying instruments and surveys to address geoscience education research (GER) questions is among the high-ranked needs in a 2016 survey of the GER community (St. John et al., 2016). The purpose of this study was to develop and validate a student-centered assessment instrument to measure course experience in a general education integrated science course. A one-shot case study of pre-experimental design (Creswell, 2009, 2014) was used to understand student experiences in a large-enrollment course with digital content integration, including out-of-class video, online presentations, and warm-up questions, as well as in-class video paired with discussion and small group activities. In two sections taught by the same geoscientist, 209 students accessed course content in an online learning management system prior to classroom instruction. We adapted the 36-item Course Experience Questionnaire (CEQ; Wilson et al., 1997) to a Web-based survey. We conducted statistical analysis on the CEQ responses, including item factor analysis, examination of communalities and measures of association, confirmatory factor analysis, and reliability and stability testing through bootstrap resampling. The statistical analyses indicate that the results in this study are comparable to those from other cultural contexts and subject areas. Given the moderate fit of the model and reasonably stable results, we propose that the core indicators of the CEQ constructs, including Appropriate Workload, Clear Goals and Standards, Generic Skills, Good Teaching, and Emphasis on Independence are sufficient to assess students' course experiences in general education science settings. These results provided moderately to strongly valid information that can help instructors in designing their technology-integrated classes, although further study with a larger sample population would be beneficial. From this study we conclude that students' perceptions of their course experiences are closely related to their development of problem-solving and analytical skills, clear course expectations to direct their own study plans, their preference for a motivating instructor, the opportunity to have a variety of learning choices, and their workload. We propose the use of a 25-item version of the CEQ that is appropriate for formative assessment in discipline-specific STEM introductory classes, including geoscience classes.
  [DOI: 10.5408/16-204.1]

Key words: course experience, assessment instrument, general education

INTRODUCTION

The learning landscape in higher education has witnessed a gradual merging of learning activities in physical classrooms with those conducted in online and/or digital environments (JISC, 2015; New Media Consortium, 2016). Assessment of students' learning experiences in these environments requires measurements that go beyond content knowledge exams (Durbin, 2002). Other assessment methods can inform teaching practice through the development and adaptation of direct and indirect measure instruments (Richardson and Price, 2003; Lo et al., 2012).

As more science courses become part of general education and faculty members explore a wide variety of teaching and learning strategies (Eagan, 2016), it is essential to understand students' learning experiences in these courses. General education science curriculum facilitates student learning of "broad knowledge of human cultures and the natural and physical world, including social sciences, science and mathematics, humanities, histories, and the arts" and "intellectual and practical skills, including effective writing, inquiry, quantitative and information literacy, and teamwork and problem solving" (Laird et al., 2009, p. 65). The research on general education science courses has indicated that students favor having access to more online instruction for learning, which allows increased focus on projects and small group work during in-class time (Baum, 2013; Swap and Walter, 2015). The teaching practice of integrating online and digital content into general education science courses demands an understanding of how students perceive their course experience. This understanding will be instrumental in developing student-centered teaching and assessment strategies to ensure quality in science, technology, engineering, and mathematics (STEM) education (President's Council of Advisors on Science and Technology [PCAST], 2012).

The results of a geoscience education research (GER) community survey in 2016 revealed that identification of instruments and surveys that would assist geoscience education researchers was a high-ranked priority among respondents (St. John et al., 2016); this disciplinary need echoes that of the broader STEM educational research community for more systematic work to identify tools and techniques useful to evaluate STEM teaching and learning (Dancy et al., 2014). Geoscience education researchers and practitioners with an interest in improving teaching and learning at the general education level have the opportunity

Received 8 April 2017; revised 27 July 2017; accepted 7 August 2017; published online 16 November 2017.

[1]Libraries and Educational Technologies, James Madison University, 1251 Carrier Drive, Harrisonburg, Virginia 22807

[2]Department of Geology and Environmental Science, James Madison University, Memorial Hall 7335, Harrisonburg, Virginia 22807

[3]Wisconsin Institute for Science Education and Community Engagement (WISCIENCE), University of Wisconsin-Madison, 445 Henry Mall, Madison, Wisconsin 53706

[a]Author to whom correspondence should be addressed. Electronic mail: Juhong.liu07@gmail.com or liujc@jmu.edu. Tel.: 540-568-2381. Fax: 540-568-7061.

to adapt assessment tools employed in other general education course settings to achieve a better understanding of their students' course experiences in discipline-specific introductory level courses. Our research aims to help address the community's need for assessment tools by formulating an approach to assess student perception of blended learning strategies in a general education integrated science course. The purpose of our study is to test the validity and reliability of a modified course experience questionnaire instrument (Wilson et al., 1997) in a non-major general education integrated science course taught by a geoscientist. We hypothesize that the validated instrument may be used in other general education STEM classes, including geosciences classes, for the following reasons: (1) the course content was multidisciplinary, drawing on physics, chemistry, Earth and space science, and human sciences; (2) the general education science class population included a mixture of majors, and (3) a range of technology-mediated instructional approaches were used.

LITERATURE CONTEXT

Need for Developing Assessment Instruments for Undergraduate Science Courses

Engaging students in introductory science classes plays an important role in undergraduates' further pursuit in STEM (PCAST, 2012). Faculty members in STEM fields have already exerted substantial efforts to optimize teaching strategies (Freeman et al., 2014; Eagan, 2016). Online and classroom-based digital technologies are being integrated into student learning in geosciences courses (Durbin, 2002; Grove, 2002; Hall-Wallace and McAuliffe, 2002; McConnell et al., 2006; Wenner et al., 2011), as they are in the context of broader undergraduate education. Students can engage with instructional content, demonstrations, and performance measures through various methods, such as video lessons and online quizzes. These extend student access to learning beyond the boundaries of the brick-and-mortar classrooms and are not limited to the scheduled in-class time with instructors. To understand how students are learning with changes in pedagogy, process, and environment, STEM education is facing a demand in developing assessment instruments (Dancy et al., 2014).

Assessment plays a major role in setting up learning expectations and outcomes, documenting the learning process, analyzing results, and improving teaching and learning practices (Angelo and Cross, 1993; Suskie, 2010). Assessment is "an approach designed to help teachers find out what students are learning . . . and how well they are learning; [It] is learner-centered, teacher-directed, mutually beneficial, formative, context-specific, ongoing, and firmly rooted in good practice" (Angelo and Cross, 1993, p. 4). Geoscience educators are exploring a wide range of assessment methods to enhance teaching and learning, but validated assessment instruments are needed to better inform and understand engaged learning for students (Durbin, 2002; Dancy et al., 2014; Renshaw, 2014; St. John et al., 2016).

Assessing Course Experience and the Course Experience Questionnaire (CEQ)

Black and William (1998) concluded that "course experience" is a core component of assessment. Course experience is a summary of students' perceptions about the quality and usefulness of instruction. Affective factors in student perception of and attitude toward teaching quality, course goals and expectations, and course workload are considered crucial components in assessing learning experiences (Libarkin, 2001; McConnell and van Der Hoeven Kraft, 2011; Jolley et al., 2012). Understanding how students perceive their learning experiences in classes needs validated instruments for measurement, comparison, and improvement (Durbin, 2002; Swap and Walter, 2015). To develop a course experience questionnaire on the basis of existing instruments, we have compared various instruments with scales that are intended to measure teaching quality, each having a specific focus (see Table I). For instance, the Attitudes and Conceptions in Science (ACS; Libarkin, 2001), Student Perceptions about Earth Sciences Survey (SPESS; Jolley et al., 2012), and Attitude toward the Subject of Chemistry Inventory (ASCI; Bauer, 2008) instruments focus on affective factors of student perception and attitude in Earth Science and chemistry university classes. In addition, McConnell and van Der Hoeven Kraft (2011) utilized Pintrich et al.'s (1991) Motivated Strategies for Learning Questionnaire (MSLQ) and discovered that student-centered teaching methods had a more positive impact on student affective perception than traditional instruction. Fraser et al. (1992) validated the Science Laboratory Environment Inventory (SLEI) instrument for measuring psychosocial factors affecting the learning environment in science laboratory classes with a large sample from universities in six different countries: USA, Canada, Australia, England, Israel, and Nigeria. Other instruments focused on the application of a particular learning theory in teaching approaches; for instance, Taylor et al. (1997) conducted an evaluative study for the urban educational system reform in Texas to promote constructivist approaches in science and mathematics education. The Constructivist Learning Environment Survey (CLES) was used to collect data in the Taylor et al. (1997) study from nearly 1,600 students in 120 grade 9–12 science classes, as well as from 494 13-year-old students in 41 science classes. The analysis results validated the instrument. The CLES was later used in a geoscience undergraduate class to measure the constructivist learning environment (Montgomery, 2005).

In a digital age, teaching infused with a variety of media and technologies is expected to enable a significant learning experience for students. Student–content interaction changes with the mediation of technologies such as digital video, web-based learning management systems, and online self-assessment. The introduction of technology results in less content dissemination directly from instructors in face-to-face settings and provides more opportunities for active learning and problem solving in classrooms. The utilization of technology tools can only be effective when aligned with strategies to engage students in mastering foundational knowledge such as information, concepts and theories, developing skills of application and critical thinking, forming the ability to integrate with a team or community, and becoming capable to perform self-directed learning (Fink, 2013). Along with the rising attention to accountability in higher education, transferable skills of students have also become a core focus of assessment (Ewell, 2009). With transferable skills including utilizing digital tools to learn and solve problems, students will be able to learn "how to learn" and apply these skills to their future learning or career. It is also crucial to measure the formulation and communication of clear goals, standards, and appropriate assessment to students. This requires the use of validated instruments to find out their learning experience in a technology-mediated course (Wiggins, 1998; Fink, 2013; Dancy et al., 2014).

The Course Experience Questionnaire (CEQ; in Table I) is intended to identify indicators of quality teaching by measuring student perception of courses and/or degree programs in higher education (Ramsden, 1991). The instrument measures six constructs: good teaching, clear goals and standards, generic skills, appropriate workload, appropriate assessment, and emphasis on independence. In addition, it has been adopted and validated in online and digital learning environments (Richardson and Price, 2003; Steele et al., 2003), and has been used to measure student course experience in STEM subjects (environmental science and computer science; Wilson et al., 1997; Richardson and Price, 2003). For these reasons, we have selected the CEQ as the instrument for our study.

In addition to being used as a national instrument to assess good teaching in Australia, the CEQ has been used and validated in a wide variety of cultural settings, including but not limited to the UK, the Netherlands, Malaysia, Japan, China, and Hong Kong (Richardson, 1994; Law and Meyer, 2011; Fryer et al., 2012; Stergiou and Airey, 2012; Yin and Wang, 2015; Thien and Ong, 2016). Course experience in classes of various subject areas were measured with this instrument (Broomfield and Bligh, 1998; Lyon and Henry, 2002; Byrne and Flood, 2003; Steele et al., 2003; Law and Meyer, 2011; Dorman, 2014; Harris and Kloubec, 2014). Researchers found that the CEQ scores were associated with student perception of learning environment, learning approaches, and learning achievement (Lizzio et al., 2002; Byrne and Flood, 2003; Richardson and Price, 2003; Diseth et al., 2006; Grace et al., 2012). The CEQ focuses on constructs that measure student perceptions of course experience and their overall satisfaction at the course or program level (Talukdar et al., 2013). The results are directly related to teachers' pedagogical practices, which can be meaningful and are easily adapted for enhancement of teaching. In the geosciences (specifically, geography), however, the CEQ data were only used for the purpose of reporting at an institutional level; not for reporting reliability or validity (Winchester, 2001).

RESEARCH QUESTION, DESIGN, AND INSTRUMENT ADAPTATION

As part of a larger assessment research project, this study was intended to contribute further data on the validity and reliability of the CEQ in the context of a technology-mediated general education integrated science class in the United States. The CEQ has been validated in various cultural and subject areas; however, the instrument was neither validated nor tested for stability in the context of a general education integrated science class in the United States. The research question of this study is: How valid and reliable is the Course Experience Questionniare for assessing students' learning experiences in general education STEM classes that are taught with digital and online components?

The original CEQ has been adapted over time. There are multiple versions (Ramsden, 1991; Wilson et al., 1997; Broomfield and Bligh, 1998), all including scales for five or six categories: Good Teaching, Clear Goals and Standards, Appropriate Workload, Appropriate Assessment, Emphasis on Independence, and Generic Skills (the optional sixth scale). Good Teaching was seen as "pitching the material at the right level, presenting it at an appropriate pace and within a clear logical structure, providing an explanation which facilitates understanding, and demonstrating both enthusiasm and empathy" (Entwistle and Tait, 1990, p. 172). Clear Goals and Standards were measured by "clear aims and objectives and clear expectations of the standard of work expected from students" (Lyon and Hendry, 2002, p. 344). Appropriate Workload measured "students' perceptions of the reasonableness of the workload" (Lyon and Hendry, 2002, p. 346). Appropriate Assessment indicated student perception of a variety of assessments measuring higher order thinking, understanding of course content, and ability of integration (Ramsden, 1991; Lyon and Hendry, 2002). Emphasis on Independence focused on student choices in the learning process (Richardson and Price, 2003). Generic Skills were defined as "problem-solving, analytic skills, teamwork, confidence in tackling unfamiliar situations, ability to plan work and written communication skills" (Wilson et al., 1997, p. 36). Because our study focused on a general education science course, general and transferrable skills such as tackling problems, analytical skills, communication, plan of work, and teamwork were important to the instructor's learning goals for students. In addition, these skills would also support students' further pursuit in the geosciences or other STEM fields, should their participation in this introductory course inspire them to pursue a STEM major. Therefore, because the 36-item version of the CEQ (Wilson et al., 1997; Table II) includes the Generic Skills scale (GS), we chose to adopt this version.

A one-shot case study of pre-experimental design (Creswell, 2009, 2014) was used to conduct this research, as illustrated in Fig. 1. The top row presents the annotation system by Creswell (2009, 2014); while the bottom row represents the design flow of this study. Group A represents the study participants, who were students from a general education science class taught by a geoscientist; X represents the course events, including student access to digital video clips, online presentations and warm-up questions, out-of-class and in-class active learning with simultaneous video watching and small group activities; O represents the measurement of the student perception of the course experience, which was the adapted CEQ from Wilson et al. (1997). In this study, an online format of the instrument was used for data collection for two reasons: (1) convenience of access to students—it was embedded in the same learning management system (LMS) where the other course materials were hosted so that students did not need a separate log-in or access point, and (2) easy data collection for researchers—the file can be downloaded directly from the LMS report as a .csv file which was compatible with Microsoft Excel and SPSS.

STUDY POPULATION AND SETTING

The research took place in a four-year university with a focus on undergraduate education, located in the mid-Atlantic region of the United States. This study focused on a general education science course called Physics, Chemistry,

TABLE I: Representative instruments for assessing student learning experience.

Published Research	Instrument	Subscales (# of items)	Purpose	Level	Subject Area
Libarkin, 2001	Attitudes and Conceptions in Science (ACS)	Attitude toward Learning Science (7)	To develop "a test aimed at assessing conception of the nature of science, as well as attitudes towards science and learning science" (p. 435)	University	General education courses in geology or planetary science, introductory geology laboratory course, introductory planetary science
		Attitude toward Science (5)			
		Conception of Science (14)			
Jolley et al., 2012	Student Perceptions about Earth Sciences Survey (SPESS)	Memorization (5)	To measure "where students lie on the novice-expert continuum, and how their perceptions change after taking one or more earth and ocean science course(s)" (p. 83)	University	Earth and ocean science
		Science and Society (3)			
		Mathematical Problem Solving (4)			
		Personal Interest (5)			
		Skeptical Reasoning (5)			
		Conceptual Problem			
		Solving (3)			
		Human-Science			
		Interaction (3)			
		Overall			
Taylor et al., 1997	Constructivist Learning Environment Survey (CLES)	Personal Relevance (6)	To provide information for monitoring the development of constructivist approaches to teaching science and mathematics by measuring student perception in science classrooms	Grades 9–12 (Taylor et al., 1997); University (Montgomery, 2005)	Middle and high school science classes; geoscience undergraduate courses
		Uncertainty of Science (6)			
		Critical Voice (6)			
		Shared Control (6)			
		Student Negotiation (6)			
Bauer, 2008	Attitude toward the Subject of Chemistry Inventory (ASCI)	Interest and Utility (5)	To measure student attitude toward learning chemistry	University	Chemistry
		Anxiety (5)			
		Intellectual Accessibility (5)			
		Fear (1)			
		Emotional Satisfaction (4)			

TABLE I: continued.

Published Research	Instrument	Subscales (# of items)	Purpose	Level	Subject Area
Pintrich et al., 1993	Motivated Strategies for Learning Questionnaire (MSLQ)	Intrinsic Goal Orientation (4)	To "present the internal consistency, reliability, and predictive validity of the MSLQ" (p. 804)	4-year, comprehensive universities	Chemistry, computer science, ecology, economics, education, English, geography/geology, history, microbiology, philosophy, physical education, psychology, sociology, Spanish
		Extrinsic Goal Orientation (4)			
		Task Value (6)			
		Control of Learning Beliefs (4)			
		Self-Efficacy for Learning and Performance (8)			
		Test Anxiety (5)			
		Cognitive and Metacognitive Strategies (CMS): Rehearsal (4)			
		CMS: Elaboration (6)			
		CMS: Organization (4)			
		CMS: Critical Thinking (5)			
		CMS: Metacognitive Self-regulation (12)			
		Resource Management Strategies (RMS): Time and Study Environment (8)			
		RMS: Effort Regulation (4)			
		RMS: Peer Learning (3)			
		RMS: Help Seeking (4)			
Fraser et al., 1992	Science Laboratory Environment Inventory (SLEI)	Student Cohesiveness (7)	To assess "perceptions of psychosocial environment in science laboratory classrooms at the higher education level" (p. 433)	Universities from six countries	Science laboratory classes
		Open-Endedness (7)			
		Integration (7)			
		Rule Clarity (7)			
		Material Environment (7)			
Wilson et al., 1997 (Adding Generic Skills scale to Ramsden, 1991, 30-item CEQ)	Modified Course Experience Questionnaire (36-item CEQ)[1]	Good Teaching (8)	To "provide further data on the construct validity [of CEQ]" (p. 36) and "to establish the reliability and validity of the new Generic Skills scale" (p. 37)	1992–1994 graduates from Australian Universities, also used in the UK, Netherlands, Malaysia, Japan, China, and Hong Kong. The items of Generic Skills are currently used in Australia at national level to assess learning experience.	Business, commerce, computing sciences, education, environmental sciences, health sciences, humanities, law, social sciences, and visual and performing arts
		Clear Goals and Standards (5)			
		Generic Skills (6)			
		Appropriate Assessment (5)			
		Appropriate Workload (7)			
		Emphasis on Independence (5)			

TABLE I: continued.

Published Research	Instrument	Subscales (# of items)	Purpose	Level	Subject Area
Entwistle et al., 2000	Approaches and Study Skill Inventory for Students (ASSIST)	Seeking Meaning (4)	To determine the relationship between student perception of learning environment in which they use certain learning approach	Universities from three countries (majority were first-year students)	Not specified
		Relating Ideas (4)			
		Use of Evidence (4)			
		Interest in Ideas (Related Subscale) (4)			
		Organized Studying (4)			
		Time Management (4)			
		Alertness to Assessment Demands (4)			
		Achieving (Related Subscale) (4)			
		Monitoring Effectiveness (Related Subscale) (4)			
		Lack of Purpose (4)			
		Unrelated Memorizing (4)			
		Syllabus-boundness (4)			
		Fear of Failure (Related Subscale) (4)			
		Supporting Understanding (4)			
		Transmitting Information (4)			
Marsh and Hocevar, 1991	Students' Evaluations of Educational Quality (SEEQ)	Instructor Enthusiasm (3)	To develop an instrument measuring teaching effectiveness	University (undergraduate and graduate courses)	Business, communication, economics, education, engineering, history, humanities, psychology, safety and systems management, social sciences, Spanish and Portuguese, system engineering
		Breadth of Coverage (2)			
		Organization/Clarity (3)			
		Assignments/Readings (2)			
		Learning/Value (2)			
		Examination/Grading (2)			
		Group Interaction (1)			
		Individual Rapport (2)			
		Workload/Difficulty (1)			
		Teacher			
		Elocutionary/Sensitivity (2)			

TABLE I: continued.

Published Research	Instrument	Subscales (# of items)	Purpose	Level	Subject Area
NSSE, FSSE, and BCSSE, 2016	National Survey of Student Engagement (NSSE)	Engagement Indicators (EI) for Theme of Academic Challenge	“To represent the multi-dimensional nature of student engagement at national, sector, institutional, and intra-institutional levels” (NSSE, FSSE, and BCSSE, 2016)	“Four-year colleges and universities about first-year and senior students’ participation in programs and activities that institutions provide for their learning and personal development” (NSSE, FSSE, and BCSSE, 2016)	
		• Higher-order Learning (4)			
		• Reflective and Integrative Learning (7)			
		• Learning Strategies (3)			
		• Quantitative Reasoning (3)			
		EI for Theme of Learning with Peers			
		• Collaborative Learning (4)			
		• Discussions with Diverse Others (4)			
		EI for Theme of Experiences with Faculty			
		• Student–Faculty Interaction (4)			
		• Effective Teaching Practices (5)			
		EI for Theme of Campus Environment			
		• Quality of Interaction (5)			
		• Supportive Environment (8)			

[1]Instrument used in this study.

TABLE II: The 36-item modified CEQ (Wilson et al., 1997). Negatively worded items are *italicized*.

Item	Item Statement	Code
Good Teaching (GT) Scale		
Q32	This course really tries to get the best out of all its students.	GT1
Q20	The instructor is extremely good at explaining things to us.	GT2
Q19	The instructor normally gives helpful feedback on how you are going.	GT3
Q3	The instructor motivates students to do their best work.	GT4
Q22	The instructor works hard to make subjects interesting.	GT5
Q29	*The instructor shows no real interest in what students have to say.*	*GT6*
Q7	The instructor put a lot of time into commenting on students' work.	GT7
Q36	The instructor makes a real effort to understand difficulties students may be having with their work.	GT8
Generic Skills (GS) Scale		
Q5	This course has sharpened my analytic skills.	GS1
Q11	This course has improved my written communication skills.	GS2
Q25	This course has helped me develop the ability to plan my own work.	GS3
Q2	This course has helped me to develop my problem-solving skills.	GS4
Q10	As a result of doing this course, I feel more confident about tackling unfamiliar problems.	GS5
Q9	This course has helped develop my ability to work as a team member.	GS6
Clear Goals and Standards (CG) Scale		
Q34	The instructor makes it clear right from the start what they expect of students.	CG1
Q6	You usually have a clear idea of where you're going and what's expected of you.	CG2
Q1	It's always easy here to know the standard of work expected.	CG3
Q16	*It's often hard to discover what's expected of you in this course.*	CG4
Q21	*The aims and objectives of this course are NOT made very clear.*	*CG5*
Appropriate Workload (AW) Scale		
Q31	*The course is overly theoretical and abstract.*	*AW1*
Q28	There was an adequate number of assignments.	AW2
Q17	We are generally given enough time to understand the things we have to learn.	AW3
Q12	*It seems to me that the syllabus tries to cover too many topics.*	*AW4*
Q24	*There's a lot of pressure on you as a student here.*	*AW5*
Q35	*The sheer volume of work to get through in this course means you can't comprehend it all thoroughly.*	*AW6*
Q4	*The workload is too heavy.*	*AW7*
Appropriate Assessment (AA) Scale		
Q30	*It would be possible to get through this course just by working hard around exam times.*	*AA1*
Q23	*Too much stuff, asks us questions just about facts.*	*AA2*
Q26	*Feedback on student work is usually provided ONLY in the form of marks and grades.*	*AA3*
Q15	*The instructor seems more interested in testing what you've memorized than what you've understood.*	*AA4*
Q8	*To do well on this course all you really need is a good memory.*	*AA5*
Emphasis on Independence (IN) Scale		
Q13	The course has encouraged me to develop my own academic interests as far as possible.	IN1
Q33	*There's very little choice in this course in the ways you are assessed.*	*IN2*
Q18	Students are given a lot of choice in the work they have to do.	IN3
Q14	Students have a great deal of choice over how they are going to learn in this course.	IN4
Q27	We often discuss with our instructor about how we are going to learn in this course.	IN5

Used with permission from the author of the CEQ, Dr. Keithia Wilson.

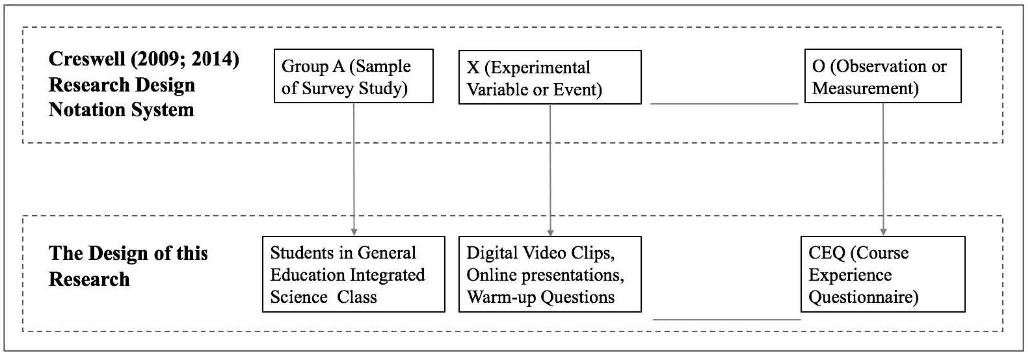

FIGURE 1: Research design model.

and the Human Experience, which is required of all nonscience majors who don't take a dedicated chemistry, physics, or astronomy class. The class is taught by faculty from all of the sciences; different sections have common goals and objectives but may cover different topics. Learning objectives focus on the nature of science and scientific literacy. In the two courses sections described in this study, content included topics such as Science versus Pseudoscience, Atoms and Bonds, Energy, Waves, Radioactivity, and Formation of Stars and Planets (Table III). Earth Science content is integrated as examples in several units when the course is taught by a geoscience instructor; for example: discussion of mineral structures in relation to atoms and chemical bonding, energy propagation and motion of ocean or seismic waves, crater counting as a means of relative age determination, and age-dating of rocks via radioactive decay.

TABLE III: Instructional content of the integrated science class in this study.

Week	Topic
1	How Science Works
2	Mechanics
3	Energy
4	Science vs. Pseudoscience
	Exam 1
5	Temperature and Heat
6	Mechanical Waves
7	Electromagnetic Waves
8	The Atom and Quantum Mechanics
	Exam 2
9	No Class: Spring Break
10	Chemical Bonding and Material Properties
11	Radioactivity
12	Particle Physics
13	Chemistry and Physics in Society
	Exam 3
14	Stars
15	Cosmology
16	Planets
Finals Week	EXAM 4

For this study, two experienced geoscience instructors paired with an instructional designer with experience in psychometric assessment methods to redesign two sections of the class by integrating digital video and online learning assessment components. The same researchers designed this research study to validate the CEQ. The research protocol was approved by the Institutional Review Board office in the university in the fall semester of 2013. In spring 2014, a total of 12 sections of the course were taught to over 750 undergraduate students at the university. One of the geoscience instructors from the research team taught two 15-week sections of the course to a total of 209 students; these two sections were the focus of this study. Students in the two sections were taught identical content and completed the same suite of learning activities and exams (Table IV). Among them, 67% were seniors, 27% juniors, 6% sophomores, and 0% freshmen.

In the two sections of this class involved in this study, the students interacted with online content through the LMS, including watching segmented video lessons and online lectures in digital formats of text or narrated presentations, completing online warm-up questions, and submitting assignments. Some in-class activities were also integrated with technologies, such as watching short video clips. There were also small in-class group activities guided by worksheets and the instructor's group visits. The classroom was equipped with an instructor's station with a computer and internet access, a remote-control projector station, and a large screen for projection. Students were required to bring a calculator capable of scientific notation to the class and were required to have daily access to the online LMS, which was available in university computer labs if the students did not have private internet access or a personal

TABLE IV: Comparison of major course activities and assessment of the two class sections.

Section A (Face-to-face)	Section B (Hybrid)
IN-CLASS: • In-Class Quizzes • In-Class Activities • Exam 1 • Exam 2 • Exam 3 • Exam 4 ONLINE: • Warm-Up Questions • Homework	IN-CLASS: • In-Class Activities • Exam 1 • Exam 2 • Exam 3 • Exam 4 ONLINE: • Warm-Up Questions • Content quizzes • Homework as blog responses

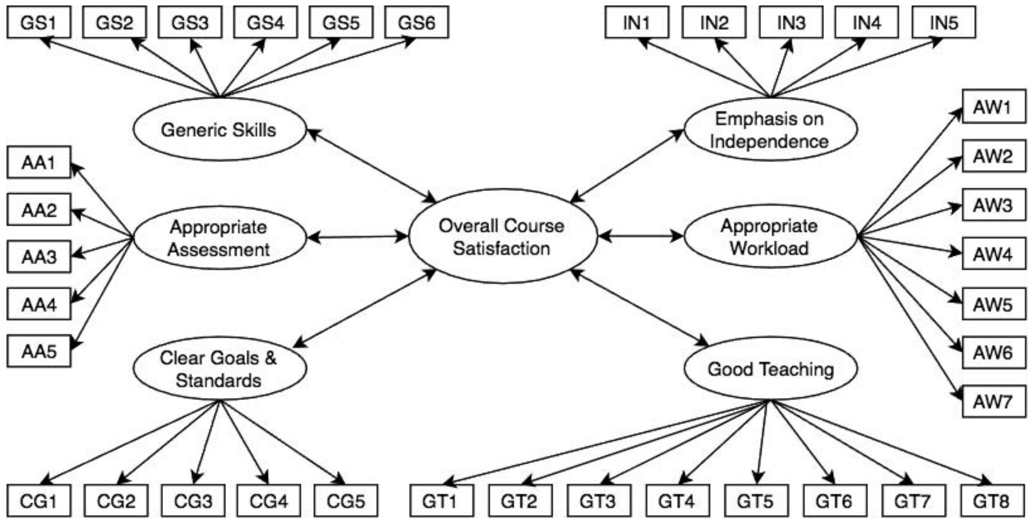

FIGURE 2: Representation of the CEQ theoretical model. The ellipses represent the six constructs in Table II. The rectangles represent the items from Table II that theoretically manifest the constructs.

computer. Students in Section A met in class twice a week, and students in Section B met online once a week and in class once a week. Aside from watching digital video instructional materials online versus in-class content delivery, the students in the two sections had very similar assessment requirements (Table IV). The small differences in assessment reflect the adjustments made for the hybrid setting of Section B.

INSTRUMENT AND DATA COLLECTION

The main instrument used in this study was the Course Experience Questionnaire (CEQ) that was adapted by Wilson et al. (1997). It consisted of 36 items organized within six scales and one Overall Satisfaction item (Table II, presented with permission from K. Wilson). Each item was associated with a 5-point Likert scale, with 1 = Strongly Disagree and 5 = Strongly Agree.

At the end of the semester, the online CEQ questionnaire and a Web-based consent form were deployed to the students in both sections. Of the 209 students, 129 gave consent to participate in the study and completed the online CEQ, yielding a response rate of 61.7%. Since the approved Institutional Review Board protocol stated that students would have the freedom to not take the survey without being subject to a penalty of any kind, not all students participated in taking the CEQ.

TABLE V: Item factor analysis results. Eigenvalue = 1.

Component	Initial Eigenvalue		
	Total	% of Variance	Cumulative %
1	9.11	25.31	25.31
2	2.80	7.77	33.08
3	2.19	6.08	39.16
4	1.93	5.36	44.52
5	1.66	4.62	49.14
6	1.42	3.96	53.10
7	1.31	3.65	56.75
8	1.26	3.49	60.24
9	1.11	3.08	63.32
10	1.07	2.96	**66.28**

DATA ANALYSIS

Historically, the CEQ has formulated a theoretical model with latent variables or constructs: Generic Skills (GS), Clear Goals and Standards (CG), Appropriate Workload (AW), Emphasis on Independence (IN), Good Teaching (GT), and indicators or items (Fig. 2), along with the CEQ development in various versions, formats, and implementation contexts (Ramsden, 1991; Wilson et al., 1997; Broomfield and Bligh, 1998). The items demonstrated varied loadings to Overall Course Satisfaction through their theoretically associated latent variables in different subject areas and contexts (Richardson, 1994; Steele et al., 2003; Fryer et al., 2012; Thien and Ong, 2016). With the 129 observed datasets, we conducted the following four major phases of data analysis with SPSS and Amos 24: (1) item factor analysis to initially confirm the original CEQ structure (Wilson et al., 1997); (2) examination of communalities and measures of association between each of the 36 items and the Overall Course Satisfaction to address the limitation of sample size; (3) confirmatory factor analysis (CFA) to estimate the instrument validity with the observed data; and (4) reliability analysis and stability testing through bootstrap resampling.

Item Factor Analysis

Item factor analysis of the datasets from 129 samples was initially conducted to confirm the CEQ structure. Ten factors accounted for the magnitude of 66.28% of the total variance (Table V), distributed as follows: Eight items loaded on Factor 1 and focused primarily on the GS scale. Six GS items along with *IN_AcademicInterests* and *GT_InstructorMotivation* accounted for 25.31% of the total variance (Table VI). Four CG items loaded on Factor 2 and accounted for 7.77% of the total variance. Three GT items and *AW_Assignments* loaded on Factor 3 and accounted for 6.07% of the total

TABLE VI: Rotated component matrix results.

	Component									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
GS_TackleProblems	0.794									
GS_AnalyticSkills	0.789									
GS_CommSkills	0.764									
GS_ProblemSolving	0.720									
IN_AcaInterests	0.711									
GS_PlanWork	0.561				0.535					
GT_InstrutorMoti	0.561									
GS_TeamMember	0.409									
CG_Exp_Standard		0.777								
CG_ClearExp		0.688								
CG_HardExpCourse		0.663								
CG_ClearIdea		0.468								
AW_Assignments			0.710							
GT_Understanding			0.649							
GT_Explain			0.589							
GT_MakeInteresting			0.517							
CG_NotClearObj[1]										
AW_Abstract				0.729						
AW_Workload				0.629						
AW_Volum				0.599						
AW_EnoughTime				0.538						
AW_ManyTopics				0.507			0.475			
IN_ChoiceLearn					0.754					
IN_ChoiceDo					0.748					
AA_MarksGrades						0.765				
GT_HelpFeedback			0.457			0.679				
GT_TimeComment						0.601				
AA_Memorized							0.692			
AA_Memory							0.692			
GT_NoInterest							0.509			
AW_Pressure								0.834		
AA_Facts								0.692		
AA_ExamTimes									0.769	
GT_AllStudents									0.480	
IN_DiscussInstr										0.746
IN_ChoiceAssess					0.408	0.436				-0.529

[1]**CG_NotClearObj** did not load on any of the 10 components.

variance. Factor 4 focused on the AW scale and five AW items accounted for 5.36% of the total variance. Factor 5 focused on the IN scale. Three IN items and *GS_PlanningWork* with split loading accounted for 4.62% of the total variance. Four items loaded on Factor 6 and accounted for 3.96% of the total variance, including two items from the GT scale, one from the original AA scale, and one from the IN scale. Three items loaded on Factor 7 and accounted for 3.65% of the total variance, with two AA items and one GT item. On Factors 8, 9, and 10, two items from different scales split loading and accounted for 3.49%, 3.08%, and 2.96% respectively of the total variance.

Examination of Communalities

The levels of item communalities have strong interaction with the sample size. "When communalities are high, good recovery of population factors can be achieved with relatively small samples" (MacCallum et al., 2001, p. 612). In social science, common magnitudes of communalities are 0.40 to 0.70 (Costello and Osborne, 2005). An item factor analysis was conducted with principal components analysis and varimax rotation in SPSS and generated the results in Table VII. The Kaser-Meyer-Olkin (KMO) measure of 0.816 indicates that it is adequate to use the factor analysis results (Meyers et al., 2006). The item communalities in Table VII are all between 0.40 and 0.80. Therefore, the sample size in this study, though small, is acceptable for further analysis.

Measures of Association

Previous CEQ studies found that some items had split loading or very weak loading on their respective constructs (Richardson, 1994; Steele et al., 2003; Fryer et al., 2012; Thien and Ong, 2016). Similar uneven loading resulted from this study (Table VI). For instance, the item "CG_NotClearObj" did not load to any of the 10 components. Statistically it means that the item of "The aims and objectives of this course are NOT made very clear" did not contribute to any of the 10 identified components.

According to the commonly accepted literature, sample sizes of 8:1 or 10:1 in the ratio between subjects to instrument items are usually desired for instrument validation tests (Mundfrom et al., 2005; Pearson and Mundform, 2010). For exploratory factor analysis with less than desired sample size, de Winter et al. (2009) recommended determining the condition of the data, with approaches such as measures of association.

Measures of association resulting from analyzing proportional reduction of error (PRE), can reflect the strength and usefulness of predicting factors to the dependent variables (Frankfort-Nachmias and Leon-Guerrero, 2010), which indicates Overall Course Satisfaction in this study. Since all the items including Overall Course Satisfaction were measured on a 5-point Likert scale, Gamma (γ) values as an ordinal measure of association were analyzed (Table VIII). A general guideline for interpreting PRE for measure of association for ordinal data is $\gamma \geq +0.2$, meaning "weak positive relationship"; $\gamma \geq +0.4$ meaning "moderate positive relationship"; $\gamma \geq +0.6$ meaning "strong positive relationship" (Frankfort-Nachmias and Leon-Guerrero, 2010). Among the 36 items, all of the AA items and items AW1, GS1, GT7, IN2, IN3, and IN5 had $\gamma < 0.4$. Therefore, the remaining 25 items with $\gamma > 0.4$ indicate moderate to strong positive relationship with the observed Overall Course Satisfaction, and these data were retained for further analysis. We refer to these 25 items as the "short form" CEQ.

Construct Validity of the CEQ—Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was conducted to further test the structure of the CEQ. "CFA permits the direct testing of the CEQ as a theoretical model through the calculation of a measure of fit between the proposed model of the designated CEQ scales, and the optimal model derived from estimates of the various relationships" (Wilson et al., 1997, p. 41). The CFA adopted structural equation modeling (SEM) goodness-of-fit indices as guides for interpretation. The analysis used maximum likelihood as the discrepancy estimation, with standardized estimates and squared multiple correlation as the output parameters.

The results of the 36-item CFA analysis revealed that $\chi^2 = 988.99$, $p < 0.001$; comparative fit index (CFI) = 0.74; goodness-of-fit index (GFI) = 0.71; normed fit index (NFI) = 0.56; 579 degrees of freedom; and root mean square error of approximation (RMSEA) = 0.07. The CFI index indicated a lack of model fit. This also supports the measures of association results that the full dataset can not be used for further analysis. Thus, we repeated these analyses with the 25-item short form version as indicated previously.

Based on the PRE (de Winter et al., 2009) and 36-item CFA analyses, the 25 items with $\gamma > 0.4$ were used to estimate the model fit using CFA (Fig. 3). These results indicated a good fit of the model, with $\chi^2 = 465.66$, $p <$

TABLE VII: Item communalities results.

	Initial	Extraction
AA1	1.000	0.624
AA2	1.000	0.658
AA3	1.000	0.719
AA4	1.000	0.587
AA5	1.000	0.531
AW1	1.000	0.634
AW2	1.000	0.654
AW3	1.000	0.739
AW4	1.000	0.697
AW5	1.000	0.726
AW6	1.000	0.560
AW7	1.000	0.663
CG1	1.000	0.725
CG2	1.000	0.584
CG3	1.000	0.681
CG4	1.000	0.727
CG5	1.000	0.490
GS1	1.000	0.771
GS2	1.000	0.709
GS3	1.000	0.737
GS4	1.000	0.679
GS5	1.000	0.818
GS6	1.000	0.548
GT1	1.000	0.669
GT2	1.000	0.716
GT3	1.000	0.760
GT4	1.000	0.618
GT5	1.000	0.569
GT6	1.000	0.579
GT7	1.000	0.667
GT8	1.000	0.625
IN1	1.000	0.600
IN2	1.000	0.764
IN3	1.000	0.649
IN4	1.000	0.660
IN5	1.000	0.720

TABLE VIII: Measures of association (γ values) of the 36 CEQ items. Bold items bear the measure of association results of γ > 0.4.

	γ of Appropriate Assessment to Overall	γ of Appropriate Workload to Overall	γ of Clear Goals and Standards to Overall	γ of Generic Skills to Overall	γ of Good Teaching to Overall	γ of Emphasis on Independence to Overall
AA1	−0.036					
AA2	0.226					
AA3	0.182					
AA4	0.282					
AA5	0.137					
AW1		**0.410**				
AW2		**0.540**				
AW3		**0.741**				
AW4		**0.538**				
AW5		0.163				
AW6		**0.612**				
AW7		**0.592**				
CG1			**0.631**			
CG2			**0.556**			
CG3			**0.443**			
CG4			**0.568**			
CG5			**0.623**			
GS1				**0.750**		
GS2				**0.527**		
GS3				**0.610**		
GS4				**0.643**		
GS5				**0.694**		
GS6				0.245		
GT1					**0.633**	
GT2					**0.699**	
GT3					**0.447**	
GT4					**0.723**	
GT5					**0.571**	
GT6					**0.426**	
GT7					0.333	
GT8					**0.606**	
IN1						**0.585**
IN2						0.261
IN3						0.326
IN4						**0.434**
IN5						0.243

0.001; CFI = 0.84; GFI = 0.78; NFI = 0.70; 265 degrees of freedom; RMSEA = 0.07; parsimony normed fit index (PNFI) = 0.62 and parsimony comparative fit index (PCFI) = 0.74. While the χ^2 was significant, the other indices were examined and interpreted to detect the model fit (Browne and Cudeck, 1992). A good fit is generally indicated by a CFI in excess of 0.8, RMSEA value between 0.05 and 0.08, and PNFI and PCFI values of 0.50 or greater. (Cole, 1987; Mulaik et al., 1989; MacCallum et al., 1996; Meyers et al., 2006). The results described above meet all of these criteria, indicating that the 25-item short form CEQ can be used as a valid instrument in this setting.

As shown in Fig. 3 and Table IX, it is evident that the 25 items load onto the five common latent variables. The GS category appears to be best represented by GS5, GS1, and GS3. Their standardized regression weights are 0.91, 0.83, and 0.74, respectively. Additionally, GS explains about 83% of the variance in GS5, 70% of the variance in GS1, and 54%

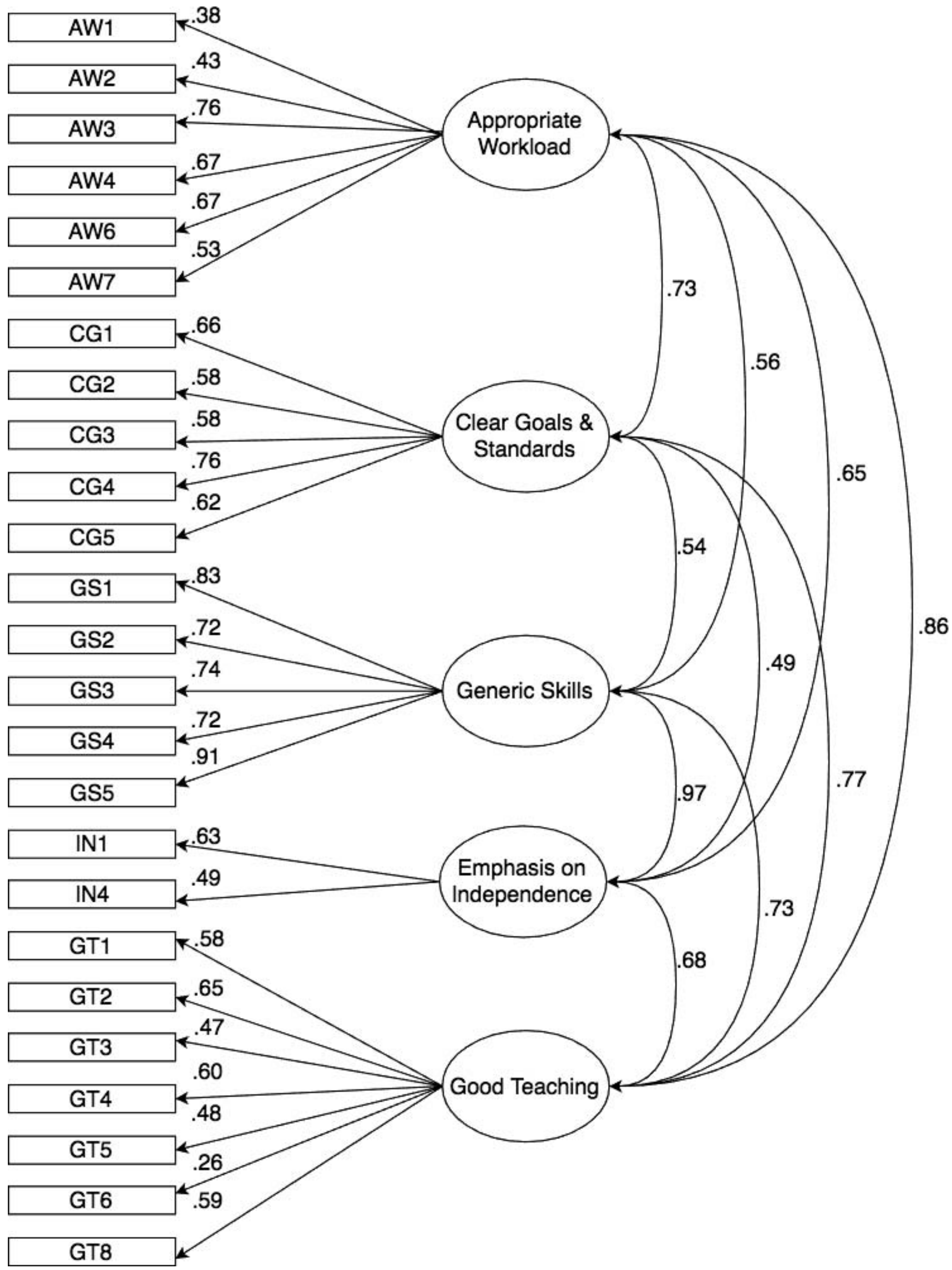

FIGURE 3: The CEQ fit model in the current study. The ellipses represent the constructs of the five scales in Table IX; the rectangles represent the observed variables (i.e., items) manifesting the constructs. The standardized regression coefficients for GS5 and GS1 are high, 0.91 and 0.83, respectively. The covariance between the construct of GS and that of IN is high, 0.97.

of the variance in GS3. The best indicators of CG appear to be CG4 and CG1. Their standardized regression weights are 0.76 and 0.66, respectively. Additionally, CG explain about 58% of the variance in CG4 and 44% of the variance in CG1. Regarding AW, AW3, AW4, and AW6 appear to be the best indicators, with standardized regression weights of 0.76, 0.67, and 0.67, respectively. AW explains 57% of the variance in AW3, 45% of the variance in both AW4 and AW6. The best indicator of IN appears to be IN1, with the standardized regression weight of 0.63, and IN explains 40% of the variance in IN1. Regarding GT, GT2, GT4 and GT8 appear to be the best indicators, with standardized regression weights of 0.65, 0.60, and 0.59, respectively. GT explains 42% of the variance in GT2, 36% of the variance in GT4, and 34% of the variance in GT8. It is worthwhile to note that GT6 bears the lowest standardized regression weight of 0.26, and GT only explains 7% of the variance in GT6, which is stated as: *GT6–*

TABLE IX: The 25-item ("short form") CEQ validated with CFA in this study. Negatively worded items are *italicized*.

Item Statement	Code	Standardized Regression Weight	Variance
Good Teaching (GT) Scale			
This course really tries to get the best out of all its students.	GT1	0.58	0.33
The instructor is extremely good at explaining things to us.	GT2	0.65	0.42
The instructor normally gives helpful feedback on how you are doing.	GT3	0.47	0.22
The instructor motivates students to do their best work.	GT4	0.60	0.36
The instructor works hard to make subjects interesting.	GT5	0.48	0.23
The instructor shows no real interest in what students have to say.	*GT6*	*0.26*	*0.07*
The instructor makes a real effort to understand difficulties students may be having with their work.	GT8	0.59	0.34
Generic Skills (GS) Scale			
This course has sharpened my analytic skills.	GS1	0.83	0.70
This course has improved my written communication skills.	GS2	0.72	0.52
This course has helped me develop the ability to plan my own work.	GS3	0.74	0.54
This course has helped me to develop my problem-solving skills.	GS4	0.72	0.52
As a result of doing this course, I feel more confident about tackling unfamiliar problems.	GS5	0.91	0.83
Clear Goals and Standards (CG) Scale			
The instructor makes it clear right from the start what they expect of students.	CG1	0.66	0.44
You usually have a clear idea of where you're going and what's expected of you.	CG2	0.58	0.34
It's always easy here to know the standard of work expected.	CG3	0.58	0.34
It's often hard to discover what's expected of you in this course.	CG4	0.76	0.58
The aims and objectives of this course are NOT made very clear.	*CG5*	*0.62*	*0.38*
Appropriate Workload (AW) Scale			
The course is overly theoretical and abstract.	*AW1*	*0.38*	*0.15*
There were an adequate number of assignments.	AW2	0.43	0.19
We are generally given enough time to understand the things we have to learn.	AW3	0.76	0.57
It seems to me that the syllabus tries to cover too many topics.	*AW4*	*0.67*	*0.45*
The sheer volume of work to be got through in this course means you can't comprehend it all thoroughly.	*AW6*	*0.67*	*0.45*
The workload is too heavy.	*AW7*	*0.53*	*0.28*
Emphasis on Independence (IN) Scale			
The course has encouraged me to develop my own academic interests as far as possible.	IN1	0.63	0.40
Students have a great deal of choice over how they are going to learn in this course.	IN4	0.49	0.24

The instructor shows no real interest in what students have to say.

Reliability and Stability

An initial test of construct reliability was performed to the 129 sets of CEQ responses (Gall et al., 2003). Prior to the data analysis, the negatively stated items shown in italicized font in Table II were reversely coded. The Cronbach's alpha coefficient was calculated to report the initial reliability. For an instrument with multiple scales like the CEQ, the Cronbach's alpha coefficient has been recommended for analysis of all scales in the instrument to demonstrate the consistency of measurement, as well as the item-total correlations (Cortina, 1993; Streiner, 2003). "Reliability coefficients vary between values of 0.00 and 1.00, with 1.00 indicating perfect reliability of the test scores and 0.00 indicating no reliability. In general, tests that yield scores with a reliability of 0.80 or higher are sufficiently reliable for most research purposes" (Gall et al., 2003, p. 196). The Cronbach's alpha coefficient for all 37 items, including 36 items in the six subscales and the Overall Course Satisfaction item was 0.905. As presented in Table X, the reliability coefficients of all six scales were analyzed and compared with the reliability tests resulted and reported from previous studies using the CEQ.

However, because "[Cronbach] α is affected by the length of the scale, high values do not guarantee internal consistency or unidimensionality" (Streiner, 2003, p. 103). To further test the consistency of dimensionality of the 25-item CEQ, a SEM analysis with bootstrap resampling was

TABLE X: Comparison of reliability coefficients of the CEQ scales.

CEQ Scales	This Study	Law and Meyer (2011)	Ramsden (1991)	Richardson (1994)
Good Teaching (GT)	0.75	0.77	0.87	0.79
Clear Goals and Standards (CG)	0.76	0.23	0.80	0.77
Generic Skills (GS)	0.87	0.78		
Appropriate Assessment (AA)	0.43	0.60	0.71	0.47
Appropriate Workload (AW)	0.71	0.55	0.77	0.71
Emphasis on Independence (IN)	0.60	0.47	0.72	0.55

conducted to provide simulated resampling for results based on the observed data. The bootstrap analysis could provide descriptive evaluation about whether the results would be reasonably stable over different reconfigurations (Thompson, 1993; Fan, 2003). The results of resampling with samples sizes of 200, 500, 1,000, and 2,000 are presented in Table XI. The estimates were stable across various simulated sample reconfigurations.

DISCUSSION

This study aimed to establish the construct validity and reliability of the Course Experience Questionnaire (CEQ) for large-enrollment general education integrated sciences classes. The interpretation of instrument validity was based on the results of SEM CFA analysis. The interpretation of reliability was based on Cronbach's alpha coefficients and the SEM bootstrapping procedure of maximum likelihood estimation. To address the limitation of the comparatively small sample size, the analysis procedure examined the item communalities and utilized PRE through analyzing ordinal measures of association (de Winter et al., 2009). The CFA analysis with the maximum likelihood model resulted in a moderate overall model fit with 25 indicators for the five latent variables in the CEQ: Generic Skills (GS), Clear Goals and Standards (CG), Appropriate Workload (AW), Emphasis on Independence (IN), and Good Teaching (GT). The results of this study have contributed reliability scores for the CEQ that are comparable to previous studies in digital and Web-based learning environments (Richardson and Price, 2003; Steele et al., 2003). The CFA results are similar to other studies that conducted similar analyses for science courses, including environmental science (Wilson et al., 1997). In this study, our 25-item short form CEQ generates a better model fit than the full 36-item CEQ.

Individual standardized regression weights for items in the 25-item CEQ are presented in Table IX. The results indicate which course aspects contribute most directly to the students' perception of a successful course experience. Factors associated with the perception of a successful course experience and their implications are discussed below.

- In the area of Generic Skills, students' perception of their confidence in tackling problems and in analytical skills were closely associated with their overall course experience. This also applies to their abilities of planning course-related work, written communication skills, and development of problem-solving skills.
- Results for the Clear Goals and Standards construct indicate that students desired to have clear expectations of them in a course and considered it important for an instructor to make those expectations clear right from the start.
- Two indicators from the original Emphasis on Independence scale resulted in the fit model of current study. Students wanted to have the maximum flexibility to develop their own academic interests and to make choices about how they were going to learn in a class. This suggests that providing material in a variety of formats may be preferred by students.
- Results for the Appropriate Workload category indicate that student overall course experience was closely related to students' perception of having enough time to understand what was learned and not having too many topics covered at one time. The volume of work, as well as workload, would also matter to student course experience.
- In the area of Good Teaching, students would perceive good teaching if an instructor was extremely good at explaining things, motivated students to do their best work for the class, and if an instructor made efforts to understand students' difficulties in coursework. Students' overall course experience was closely associated to the fact that an instructor tried to get the best out of all students, worked hard to make the course interesting, and provided helpful feedback.

These results provided moderate to strongly valid information to help instructors in designing technology-integrated classes. In the two course sections, the sample population perceived their course experience as most closely related to: (1) their development of problem-solving and analytical skills, (2) clear course expectations to direct their own study plans, (3) their preference for a motivating instructor, (4) the opportunity to have a variety of learning choices, and (5) an appropriate workload. We infer that with increasing digital dissemination of course content, the design focus of general education science courses ought to be on active learning strategies that nurture student scientific inquiries and maximize their choices to develop cognitive capabilities, rather than loading students with intense workload and significant course content requirements. These results are similar to previous CEQ studies that were conducted to examine the relationship between student course experiences, deep, strategic, and surface learning approaches, and academic achievements in classroom-based courses (Entwistle and Tait, 1990; Diseth, 2007). These will be meaningful to investigate further as student learning expands in digitally equipped learning environments in order to determine how instruction and learning environments can be designed to optimize students' learning experiences. The results also help set up the linkage between

TABLE XI: Bootstrapping procedure results for sample sizes of 200, 500, 1,000, and 2,000.[1]

	Estimate				SE				Mean				Lower–Upper			
	200	500	1,000	2,000	200	500	1,000	2,000	200	500	1,000	2,000	200	500	1,000	2,000
Appropriate Workload	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.20	0.20	0.20	0.20	0.05–0.44	0.05–0.43	0.05–0.43	0.05–0.41
Clear Goals and Standards	0.25	0.25	0.25	0.25	0.07	0.07	0.07	0.07	0.26	0.25	0.26	0.26	0.14–0.39	0.15–0.39	0.15–0.39	0.15–0.38
Generic Skills	0.68	0.68	0.68	0.68	0.13	0.12	0.12	0.12	0.68	0.68	0.68	0.68	0.45–0.89	0.47–0.88	0.49–0.88	0.48–0.88
Emphasis on Independence	0.46	0.46	0.46	0.46	0.19	0.20	0.18	0.18	0.48	0.48	0.48	0.47	0.22–0.78	0.21–0.77	0.22–0.78	0.21–0.80
Good Teaching	0.26	0.26	0.26	0.26	0.09	0.09	0.10	0.10	0.25	0.25	0.26	0.25	0.10–0.41	0.10–0.42	0.10–0.43	0.10–0.42

[1]Parameter: Bootstrap with maximum likelihood; confidence interval = 0.90; $p < 0.10$.

learning activities in introductory science classes to student development of transferable skills for future learning and career paths in STEM areas. For introductory geoscience classes in particular, incorporating instructional strategies that students perceive as contributing positively to their overall course experience across lecture, lab, and field learning activities may be particularly impactful.

LIMITATIONS

The study was designed to contribute to research on assessment of learning. The results are meaningful to the instructional design and learning assessment for general education science classes. However, the study has three default limitations: (1) lack of demographic information, (2) convenience sampling vs. random sampling, and (3) small sample size. The first limitation is the missed demographic information of the population. At the time of research design, the researchers received support from the institution to provide student demographic information through the university's registrar system. Immediately at the completion of the semester, however, the instructor who taught the two sections shifted to a new job in a different university. Consequently, the researchers were not able to access and incorporate the demographic information. The second limitation is the convenience sampling, which seems prevalent in educational research. The sample was not representative enough to assure generalizable results. Furthermore, the stability test results through simulation with the bootstrapping procedure were not based on empirical data, so they may have generated overly optimistic values (de Winter et al., 2009). The third limitation is the small sample size. While tests were applied to demonstrate the reliability of the small sample size and a moderately fit model resulted, we also accept the caveat that "the resulting model and solution are not an exact or complete representation of phenomena operating in the real world, but only a parsimonious approximation" (MacCallum et al., 2001, p. 613). Thus, the results should be further validated with larger sample sizes, despite the stable results generated through the bootstrap resampling procedure (Fan, 2003).

In general, the limitations caution against the generalizability of the results to different types and levels of geoscience classes. As a case study at a single university, our study fits into the second level of the GER strength of evidence pyramid (St. John and McNeal, 2017). Higher confidence in the generalizability of findings would come from additional testing in a multi-institutional study of geosciences courses that includes students from different demographic groups.

IMPLICATIONS FOR GEOSCIENCE INSTRUCTION AND GEOSCIENCE EDUCATION RESEARCH

Students' perception of their course experience is influenced by the skills they feel they have gained and faculty attention to course design. Transferrable skills through clear explanation, instructors' motivating techniques, and helpful comments are important to students' positive course experience. Pedagogical practice in science education has indicated that spending in-class time with active learning generated positive learning experiences, though this may need more research in the geosciences according to a recent meta-analysis study (Freeman et al., 2014).

A core instructional implication resulting from our study is that the structure of both out-of-class and in-class activities needs to align with clearly articulated course goals, learning objectives/achievement standards, and course workload. This may require intentional collaboration in course design among instructors, instructional designers, and students. We conclude that faculty in general education science courses could use the newly modified 25-item CEQ (Table IX) as a means of formative assessment of student perceptions of teaching. The findings would aid continual improvement of teaching as well as contribute to the further development of this instrument.

Implications for geoscience education research center on assessment. The results of this study challenge further research on the wording of constructs and items that measure appropriate assessment in student perception. Many items in the modified CEQ were negatively worded, and further investigation into alternative wording of the question stems may be necessary (Barnette, 2000). The limitations of this study also imply the potential of further studies. The suitability of the modified instrument should be further validated with various populations and larger sample sizes.

At a broader scale, other studies that can evaluate assessment instruments are needed to assist the improvement of teaching and learning in all STEM classes, including geoscience classes. While the contexts of many general education geoscience courses are likely very similar to the general education integrated science courses in this study, the generalizability to geoscience courses with different types of teaching technologies and with laboratory or field work components, especially at the upper level, would require additional study. Assessment methods for introductory sciences and geoscience classes taught with different techniques and having different structure need development. For instance, how do laboratory-based and/or field-based geoscience classes affect student course experience? Items from the Science Laboratory Environment Inventory (SLEI; Fraser et al., 1992), those from the theme of Academic Challenge in NSSE (Table I), and the validated items from the CEQ used in this current study may be the pool for a potential assessment instrument. Face validity and content validity can be established by inviting raters among geoscience education experts (Cronbach and Meehl, 1955; Nevo, 1985; Rubio, et al., 2003). To enable generalization of these newly developed assessment instruments to upper-level, field-based, and/or lab-based geoscience classes, additional studies resulting from collaboration of geoscience educators, geoscience researchers, and researchers with expertise in psychometric assessment methods is crucial. This will make it possible to generate results for different course levels, content concentrations, and student populations.

Acknowledgments

This research would not have been possible without the support from a General Education Grant at James Madison University and the persistent collaboration among the researchers from different discipline areas. The authors want to extend collegial appreciation to Dr. Keithia Wilson from Griffith University for her permission to reuse the 37-item Course Experience Questionnaire in this study. The authors also want to acknowledge Dr. Julie Libarkin from Michigan State University and two anonymous reviewers for their constructive comments and suggestions on the manuscript and for insights on data analysis methods.

REFERENCES

Angelo, T.A., and Cross, K.P. 1993. Classroom assessment techniques: A handbook for college teachers. San Francisco, CA: Jossey-Bass.

Barnette, J.J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3):361–370.

Bauer, C.F. 2008. Attitude toward chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of Chemical Education*, 85(10):1440–1445.

Baum, Edward J. 2013. Augmenting guided-inquiry learning with a blended classroom approach. *Journal of College Science Teaching*, 42(6):27–33.

Black, P., and William, D. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1):7–74.

Broomfield, D., and Bligh, J. 1998. An evaluation of the short form course experience questionnaire with medical students. *Medical Education*, 32(4):367–369.

Browne, M.W., and Cudeck, R. 1992. Alternative ways of assessing model fit. *Sociological Methods and Research*, 21(2):230–258.

Byrne, M., and Flood, B. 2003. Assessing the teaching quality of accounting programmes: An evaluation of the Course Experience Questionnaire. *Assessment and Evaluation in Higher Education*, 28(2):135–145.

Cole, D.A. 1987. Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, 55(4):584–594.

Cortina, J.M. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1):98.

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7):1–9.

Creswell, J.W. 2009. Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J.W. 2014. Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Thousand Oaks, CA: Sage.

Cronbach, L.J., and Meehl, P.E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.

Dancy, M., Hora, M., Ferrare, J., Iverson, E., Lattuca, L., and Turns, J., 2014. Describing and measuring undergraduate STEM teaching practice: A report from a national meeting on the measurement of undergraduate science, technology, engineering and mathematics (STEM) teaching. American Association for the Advancement of Science. 57 p.

de Winter, J.D., Dodou, D., and Wieringa, P.A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44(2):147–181.

Diseth, Å. 2007. Approaches to learning, course experience and examination grade among undergraduate psychology students: Testing of mediator effects and construct validity. *Studies in Higher Education*, 32(3):373–388.

Diseth, Å., Pallesen, S., Hovland, A., and Larsen, S. 2006. Course experience, approaches to learning and academic achievement. *Education + Training*, 48(2-3):156–169.

Dorman, J.P. 2014. Classroom psychosocial environment and course experiences in pre-service teacher education courses at an Australian university. *Studies in Higher Education*, 39(1):34–47.

Durbin, J.M. 2002. The benefits of combining computer technology and traditional teaching methods in large enrollment geoscience classes. *Journal of Geoscience Education*, 50(1):56–63.

Eagan, K. 2016. Becoming more student-centered? An examination of faculty teaching practices across STEM and non-STEM disciplines between 2004 and 2014. A report prepared for the Alfred P. Sloan Foundation.

Entwistle, N., and Tait, H. 1990. Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education*, *19*(2):169–194.

Entwistle, N., Tait, H., and McCune, V. 2000. Patterns of response to an approaches to studying inventory across contrasting groups and contexts. *European Journal of Psychology of Education*, 15(1):33–48.

Ewell, P. 2009. Assessment, accountability, and improvement: Revisiting the tension. National Institute for Learning Outcomes Assessment.

Fan, X. 2003. Using commonly available software for bootstrapping in both substantive and measurement analyses. *Educational and Psychological Measurement*, 63(1):24–50.

Fink, L.D. 2013. Creating significant learning experiences: An integrated approach to designing college courses. San Francisco, CA: Jossey-Bass.

Frankfort-Nachmias, C., and Leon-Guerrero, A. 2010. Social statistics for a diverse society. Thousand Oaks, CA: Sage.

Fraser, B.J., Giddings, G.J., and McRobbie, C.J. 1992. Assessment of the psychosocial environment of university science laboratory classrooms: A cross-national study. *Higher Education,* 24(4):431–451.

Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., and Wenderoth, M.P., 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, USA*, 111(23):8410–8415, doi: 10.1073/pnas.1319030111.

Fryer, L.K., Ginns, P., Walker, R.A., and Nakao, K. 2012. The adaptation and validation of the CEQ and the R-SPQ-2F to the Japanese tertiary environment. *British Journal of Educational Psychology*, 82(4):549–563.

Gall, M.D., Gall, J.P. and Borg, W.R., 2002. Educational research: An introduction (7th ed.). Boston, MA: Allyn & Bacon.

Grace, D., Weaven, S., Bodey, K., Ross, M., and Weaven, K. 2012. Putting student evaluations into perspective: The course experience quality and satisfaction model (CEQS). *Studies in Educational Evaluation*, 38(2):35–43.

Grove, K. 2002. Using online homework to engage students in a geoscience course for general education. *Journal of Geoscience Education*, 50(5):566–574.

Hall-Wallace, M.K., and McAuliffe, C.M. 2002. Design, implementation, and evaluation of GIS-based learning materials in an introductory geoscience course. *Journal of Geoscience Education*, 50(1):5–14.

Harris, C., and Kloubec, J. 2014. Assessment of student experience in a problem-based learning course using the course experience questionnaire. *Journal of Nutrition Education and Behavior,* 46(4):315–319.

Joint Information Systems Committee (JISC). 2016. Changing the learning landscape. Available at https://www.jisc.ac.uk/rd/projects/changing-the-learning-landscape (accessed 13 August 2017).

Jolley, A., Lane, E., Kennedy, B., and Frappé-Sénéclauze, T.P. 2012. SPESS: A new instrument for measuring student perceptions in Earth and ocean science. *Journal of Geoscience Education*, 60(1):83–91.

Laird, T.F.N., Niskode-Dossett, A.S., and Kuh, G.D. 2009. What general education courses contribute to essential learning outcomes? *Journal of General Education*, 58(2):65–84.

Law, D.C., and Meyer, J.H. 2011. Adaptation and validation of the course experience questionnaire in the context of post-secondary education in Hong Kong. *Quality Assurance in Education*, 19(1):50–66.

Libarkin, J.C. 2001. Development of an assessment of student conception of the nature of science. *Journal of Geoscience Education*, 49(5):435–442.

Lizzio, A., Wilson, K., and Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education*, 27(1):27–52.

Lo, C.C., Johnson, E.L., and Tenorio, K.A. 2012. Promoting student learning by having college students participate in an online environment. *Journal of the Scholarship of Teaching and Learning*, 11(2):1–15.

Lyon, P.M., and Hendry, G.D. 2002. The use of the Course Experience Questionnaire as a monitoring evaluation tool in a problem-based medical programme. *Assessment and Evaluation in Higher Education*, 27(4):339–352.

MacCallum, R.C., Browne, M.W., and Sugawara, H.M. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2):130–149.

MacCallum, R.C., Widaman, K.F., Preacher, K.J., and Hong, S. 2001. Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36(4):611–637.

Marsh, H.W., and Hocevar, D. 1991. The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education*, 7(1):9–18.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H., Greer, L., and Heaney, P. 2006. Using concept tests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, 54(1):61–68.

McConnell, D.A., and van Der Hoeven Kraft, K.J. 2011. Affective domain and student learning in the geosciences. *Journal of Geoscience Education*, 59(3):106–110.

Meyers, L.S., Gamst, G., and Guarino, A. J. 2006. Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

Montgomery, H. 2005. Constructivist pedagogical preferences of undergraduate students in geoscience service courses at the University of Texas at Dallas. *Texas Science Teacher*, 34(1):25–30.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., and Stilwell, C.D. 1989. Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3):430–445.

Mundfrom, D.J., Shaw, D.G. and Ke, T.L. 2005. Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2):159–168.

National Survey of Student Engagement (NSSE), Faculty Survey of Student Engagement (FSSE), and Beginning College Survey of Student Engagement (BCSSE). 2016. National survey of student engagement. Available at http://nsse.indiana.edu/ (accessed 13 August 2017).

Nevo, B. 1985. Face validity revisited. *Journal of Educational Measurement*, 22(4):287–293.

New Media Consortium. 2016. Horizon report: 2016 higher education edition. Available at http://www.nmc.org/publication/nmc-horizon-report-2016-higher-education-edition/ (accessed 13 August 2017).

Pearson, R.H. and Mundform, D.J. 2010. Recommended sample size for conducting exploratory factor analysis on dichotomous data. *Journal of Modern Applied Statistical Methods*, 9(2):359–368.

Pintrich, P.R., Smith, D.A., García, T., and McKeachie, W.J. 1993. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3):801–813.

President's Council of Advisors on Science and Technology (PCAST). 2012. Report to the President: Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf (accessed 21 August 2017).

Ramsden, P. 1991. A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16(2):129–150.

Renshaw, C.E. 2014. Design and assessment of a skills-based geoscience curriculum. *Journal of Geoscience Education*, 62(4):668–678. doi:10.5408/13-100.1

Richardson, J.T. 1994. A British evaluation of the course experience questionnaire. *Studies in Higher Education*, 19(1):59–68.

Richardson, J.T., and Price, L. 2003. Approaches to studying and perceptions of academic quality in electronically delivered courses. *British Journal of Educational Technology*, 34(1):45–56.

Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S., and Rauch, S. 2003. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2):94–104.

Steele, G.A., West, S.A., and Simeon, D.T. 2003. Using a modified Course Experience Questionnaire (CEQ) to evaluate the innovative teaching of medical communication skills. *Education for Health: Change in Learning and Practice*. 16(2):133–144.

St. John, K., and McNeal, K. (2017). The strength of evidence pyramid: One approach for characterizing the strength of evidence of geoscience education research (GER) community claims. *Journal of Geoscience Education*, 65(4):363–372.

St. John, K., McNeal, K., Macdonald, H., and Kastens, K. 2016. Webinar: Results from the geoscience education research (GER) survey—Data for setting priorities in the GER community. Available at: http://nagt.org/nagt/profdev/workshops/GER_community/GER_webinar.html (accessed 13 August 2017).

Stergiou, D.P., and Airey, D. 2012. Using the Course Experience Questionnaire for evaluating undergraduate tourism management courses in Greece. *Journal of Hospitality, Leisure, Sport and Tourism Education*, 11(1):41–49.

Streiner, D.L. 2003. Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1):99–103.

Suskie, L. 2010. Assessing student learning: A common sense guide. New York: John Wiley & Sons.

Swap, R.J., and Walter, J.A. 2015. An approach to engaging students in a large-enrollment, introductory STEM college course. *Journal of the Scholarship of Teaching and Learning*, 15(5):1–21.

Talukdar, J., Aspland, T., and Datta, P. 2013. Australian higher education and the Course Experience Questionnaire: Insights, implications, and recommendations. *Australian Universities' Review*, 55(1):27–35.

Taylor, P.C, Fraser, B.J., and Fisher, D.L. 1997. Monitoring constructivist classroom learning environments. *International Journal of Educational Research*. 27(4):293–302.

Thien, L.M., and Ong, M.Y. 2016. The applicability of course experience questionnaire for a Malaysian university context. *Quality Assurance in Education*, 24(1):41–55.

Thompson, B. 1993. The use of statistical significance tests in research: Bootstrap and other alternatives. *The Journal of Experimental Education*, 61(4):361–377.

Wenner, J.M., Burn, H.E., and Baer, E.M. 2011. The math you need, when you need it: Online modules that remediate mathematical skills in introductory geoscience courses. *Journal of College Science Teaching*, 41(1):16–24.

Wiggins, G.P. 1998. Educative assessment: Designing assessments to inform and improve student performance (vol. 1). San Francisco, CA: Jossey-Bass.

Wilson, K.L., Lizzio, A., and Ramsden, P. 1997. The development, validation, and application of the Course Experience Questionnaire. *Studies in Higher Education*, 22(1):33–53.

Winchester, H.P. 2001. The relationship between teaching and research in Australian geography. *Journal of Geography in Higher Education*, 25(1):117–120.

Yin, H., and Wang, W. 2015. Assessing and improving the quality of undergraduate teaching in China: the Course Experience Questionnaire. *Assessment & Evaluation in Higher Education*, 40(8):1032–1049.