Work-in-Progress Paper: Creating and Validating the Conceptual Assessment for Sedimentology courses

Anna Stepanova
Department of Computer Science & Engineering
Texas A&M University
College Station, TX, USA
anna_stepanova@tamu.edu

Saira Anwar
Department of Multidisciplinary Engineering
Texas A&M University
College Station, TX, USA
sairaanwar@tamu.edu

Juan Carlos Laya
Department of Geology & Geophysics
Texas A&M University
College Station, TX, USA
layajc@tamu.edu

Carlos A. Alvarez Zarikian
International Ocean Discovery Program
Texas A&M University
College Station, TX, USA
zarikian@iodp.tamu.edu

Tracy Hammond
Department of Computer Science & Engineering
Texas A&M University
College Station, TX, USA
hammond@tamu.edu

2024 IEEE Frontiers in Education Conference (FIE) | 979-8-3503-5150-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/FIE61694.2024.10893537

***Abstract*—Contribution: In this work-in-progress paper we describe the process of creating and validating a conceptual assessment in the field of sedimentology for undergraduate geoscience courses. The mechanism can aid future geoscience educators and researchers in the process of academic assessment development aligned with learning objectives in these courses.**

Background: Prior literature review supports the benefits of using active learning tools in STEM (Science, Technology, Engineering, and Mathematics) courses. This paper is part of a larger project to develop and incorporate research-based active learning software in sedimentology and other geoscience courses to improve grade point average (GPA) and time to graduation for Hispanic students at Texas A&M University. To evaluate the novel tool, we designed and validated the conceptual assessment instrument presented in this work.

Research Question: What is the process to develop and validate a conceptual assessment for sedimentology?

Methodology: This paper follows quantitative analysis and the assessment triangle approach and focuses on cognition, observation, and interpretation to design and evaluate the conceptual assessment. In the cognition element of the triangle, we explain the mechanism for creating the assessment instrument using students' learning objectives. The observation element explains the mechanism of data collection and the instrument revision. The interpretation element explains the results of the validation process using item response theory and reliability measures. We collected the conceptual assessment data from 17 participants enrolled in two courses where sedimentology topics are taught. Participants were geology majors in one of the courses and engineering majors in the other.

Findings: The team developed a conceptual assessment that included eight multiple-choice (MCQ) and four open-ended response questions. The results of the design process described the conceptualization of questions and their validation. Also, the validity of created rubrics was established using inter-rater reliability measures, which showed good agreement between raters. Additionally, the results of the validation process indicated that the conceptual assessment was designed for students with average abilities.

***Index Terms*—geoscience education, conceptual assessment**

I. INTRODUCTION

Sedimentology is a branch of geology that studies sedimentary rocks and the related processes. These processes occur at the earth's surface and shallow subsurface and are directly related to environmental changes through geological time [1]. Sedimentology skills include rock type identification, textural analysis and sedimentary structures recognition in order to conduct detailed sedimentological investigations, and are key to understanding earth processes and environmental change.

Recently, Texas A&M University was awarded the status of an HSI (Hispanic Serving Institute) in the southeastern region of the USA due to its large number of Hispanic students. Hispanic students represent 25% in undergraduate geoscience courses and 25-30% of the First Time in College students. The Hispanic student performance indicators are markedly different from those of other ethnicities, with Hispanic students consistently having lower GPAs at graduation compared to all other ethnicities. To improve sedimentology teaching at Texas A&M University HSI, we proposed to develop SedimentSketch, an active learning educational software on sedimentology that will provide students with automatic feedback, allow sketching, and remote access to lab materials for online learning [2]. It was previously shown that active learning tools and remote learning can improve grades for Hispanic students [3], [4]. SedimentSketch will support active learning and allow students to practice from home, receive individual feedback, and improve their situational interest in geoscience as they progress in learning sedimentology.

Our long-term goals are to develop and evaluate the effectiveness of the SedimentSketch application for teaching sedimentology. In order to assess learning before and after the use of SedimentSketch, we developed a conceptual assessment for Sedimentology. The 12-question conceptual assessment was developed to assess students' knowledge of material covered in the Sedimentology course.

A. Research Question

What is the development and validation process for conceptual assessment for sedimentology?

II. Related Work

Active learning in STEM and geosciences has several advantages over traditional educational formats for in-person and online environments [5], [6]. It was shown to increase students' performance compared to lectures across disciplines [7]. Additionally, active learning activities help build learner participation and facilitate retention of the material. Active learning strategies result in overall improvements in students' achievements on exams, retention in courses, and logical thinking skills [5], [8]. Literature supports the use of active learning tools, specifically sketch-based applications that provide immediate feedback and provide an additional practice tool to support active learning. They can improve engagement, knowledge, and motivation compared to traditional learning environments. We proposed to develop the SedimentSketch application, that will support active learning and facilitate student mastery of sedimentology skills and provide a unique interactive environment for students to practice and receive individual feedback as they progress.

Concept tests have been developed for many science disciplines, to be used as measures of student learning [9]. One of the existing tools for assessment in entry-level geoscience courses is the Geoscience Concept Inventory (GCI) [10]. It represents a bank of 69 validated questions with learning being evaluated using pre and post average scores, and tested in over 42 institutions nationwide. The questions could be selected by an instructor to create a customized subtest for use in their course. Besides GCI, a number of theme-specific or geoscience subject-specific assessments were developed. In another study the researchers developed the Landscape Identification and Formation Test, a concept inventory for measuring abilities to identify landscapes and their formation timespans [11]. A statistically validated mineralogy concept assessment was developed by Scribner and Harris [12]. Another statistically validated test was developed to assess concept knowledge on a geological time scale for undergraduate students [13]. Another 21-item validated test to assess climate change understanding was developed by Libarkin and co-authors [14]. Although many of these assessment instruments contain questions that can be used for assessments in Sedimentology, no validated sedimentology-specific assessments exist.

III. Methods

A. Research Design

In this paper, the research design follows the standardized method of assessment triangle [15] and quantitative approach. For the assessment triangle method, prior literature describes three major aspects to evaluate and assess instruments, including assessment instruments, which are: 1) Cognition (theory of understanding for learning), 2) Observation (tasks that are performed), and 3) Interpretation (methods and tools for meaning-making). This model allows for establishing connections between different components of an assessment system. Following the guidelines explained in [15], [16], this paper used active learning principles, learning objectives, and expert knowledge to create the instrument and its rubrics (cognition). We pilot-tested the conceptual assessment and used a quantitative approach to establish inter-rater reliabilities, revised rubrics and assessment (observation), and used Item response theory and reliability measures to validate the assessment (interpretation).

B. Participants

We collected the pre and post assessment data from 17 students at Texas A&M University. Participants were students studying sedimentology in two different courses, in one of the courses students were geology majors (Sedimentology & Stratigraphy) and in the other, engineering majors (Geology of Petroleum). We follow the approved study protocol IRB2023-1037M, with an expiration date of 11/28/2026. The assessment was conducted at the beginning of the class and students had up to 15 minutes to complete it.

C. Data collection

The conceptual assessment was designed to assess students learning before and after they learned sedimentology in traditional class, and later with the use of SedimentSketch. For the validation process, we conducted conceptual assessments at two-time points in the traditional instruction of two geoscience courses: Sedimentology & Stratigraphy and Geology of Petroleum. The instructors in the two courses administered the pre-assessment a week before instruction began, and the post-assessment at the end of the semester.

D. Design of the conceptual assessment

For the cognition corner of the assessment triangle, sedimentology expert and two geologists (non-sedimentology specialists) developed the initial conceptual assessment. They developed a set of questions to identify key concepts in sedimentology courses to assess their knowledge of the basic sedimentology terms, simple routine steps to identify sedimentary rocks, main applications of sedimentology for climate and sea-level change reconstruction, and industry applications. The initial conceptual assessment included 12 questions (multiple choice and open-ended questions, see Table I). These questions were developed to test students' knowledge of the material covered in the Sedimentology course and SedimentSketch application. We included a mix of low-difficulty and high-difficulty questions. The content of the conceptual assessment aligns with the courses' sedimentology modules' learning objectives as identified in the course curriculum.

TABLE I
CONCEPTUAL ASSESSMENT, EXAMPLES OF QUESTIONS. QUESTIONS 1-4 ARE MCQ, AND 9-10 ARE OPEN-ENDED

Q1: What are the three main types of sedimentary rocks?
Q2: What sedimentary structures can you see in a core?
Q3: What are the steps to describe a sedimentary rock?
Q4: The angularity of grains in a sedimentary rock is referred to as:
Q5: Which statement best describes a pelagic sediment?
Q6: Main grains found in an Arkose are:
Q7: Lamina is thicker than bed. True or False.
Q8: Which one of the following features is NOT associated with sedimentary rocks?
Q9: In your own words define the term "sedimentary facies".
Q10: How sedimentary rocks can be used to reconstruct sea-level changes?
Q11: What are the processes/mechanisms involved in sedimentary rocks formation? Name all six.
Q12: What tasks would sedimentologists work on in energy industry, water resources, and climate change research? Name at least 2.
Question that was removed: Explain how climate can be inferred from sedimentary deposits.

A rubric was designed to grade MCQs and each of the open-ended questions by a team of engineering education and geoscience experts. For multiple choice questions 0 corresponded to an incorrect response, and 1 for a correct response. The rubric included a partial grade option for the open-ended questions. The students scored 0 for an incorrect response, 1 for a partially correct response, and 2 for a correct response. Since the conceptual assessment was designed keeping in mind the current curriculum and future implementation of the SedimentSketch tool, the questions were designed for the medium difficulty level and ensured that the content of each question was covered in courses teaching sedimentology with or without the use of SedimentSketch application.

E. Revision of the conceptual assessment

For the observation corner of the assessment triangle, we first focused on revising questions and associated rubrics. We used the pre-assessment responses to refine the rubrics for each question based on the inter-rater reliability and expert discussions. One of the open-ended questions was flagged during the rubrics revision process. The initial rubric flagged issues in grading at the pre-assessment step. The reliability measures flagged the same issue, so this question was dropped from the remaining validation process.

The revised conceptual assessment included (11 questions, which were identical in both pre-and post-assessment). In addition, a new open-ended question and its associated rubric were introduced in the post-assessment. The resultant conceptual assessment for validation includes 12 questions: 8 multiple choice questions, 3 open-ended questions (administered in both pre and post-assessment), and 1 open-ended question (administered in post-assessment only, see Table I).

F. Data analysis

For the interpretation corner of the assessment triangle, we completed a series of steps to establish the validity and reliability of the conceptual assessment. First, we piloted the conceptual assessment in two geoscience classes in a pre- and post-manner. We created each question's data set by combining pre- and post-assessment responses for the validation process. We used these assessment responses to establish the inter-rater reliability between the manual graders of open-ended responses using IBM SPSS Statistics 29.0.1.1 version [17]. Cohen Kappa, percent agreement, and Pearson correlation were used to measure the inter-rater reliability. The minimum criteria of 0.65 of the Cohen Kappa value (in the range of 0.60 to 0.80) was set for inter-rater reliability, indicating the benchmark for a substantial and good agreement [18]–[21]. Lastly, in addition to inter-rater reliability, we conducted a series of reliability measures and validation processes. We used item response theory methods [22], [23] to analyze item functions for reliability and validation purposes using the Rasch model and packages, including TAM [24], WrightMap [25], and tidyverse [26]. We conducted the analysis using Rstudio. We used the Expected A Posteriori (EAP) reliability, Warm's Mean Weighted Likelihood Estimation (WLE) reliability, and Cronbach alpha for assessment reliability.

IV. RESULTS

As all the open-ended questions on conceptual assessment were manually graded, we calculated inter-rater reliability using Cohen Kappa (k) statistics. Three graders graded the open-ended assessment questions to establish the inter-rater reliability for all pre- and post- student assessment data. In the analysis, one question appeared problematic in establishing reliability, and the experts removed the question from the conceptual assessment. The Cohen Kappa agreement between the graders ranged from 0.685 to 0.896 (other than the problematic question where $k = 0.336$), which shows a significant, and good agreement between raters [19], [20]. In addition to Cohen's kappa statistics, we calculated the Pearson correlation coefficient between graders, and the results ranged from 0.851 to 0.979 (except for the problematic question where $r < 0.5$). We also calculated the percent agreement between raters, which ranged from 65% to 94.12% (problematic question = 52.94%), indicating a good agreement between graders. Based on different inter-rater reliability metrics, our inter-rater reliability analyses indicated substantial agreements between raters.

Multiple reliability measures were obtained after removing the problematic question, and demonstrated similar results. For example, the Expected A Posteriori (EAP) reliability was 0.602, with a variance of 0.696. The Warm's Mean Weighted Likelihood Estimation (WLE) reliability was 0.547, and Cronbach's alpha was 0.821. These reliability values indicated good data estimates and the questions' internal consistency and responses [27], [28]. We used question difficulty level with the criteria of zero describing average ability, greater than zero for a question designed for students with higher ability,

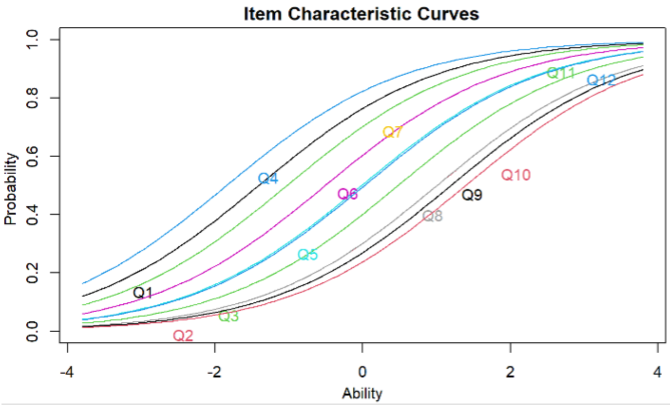

Fig. 1. Questions characteristics curves reflecting questions' difficulty level, based on students' ability and probability

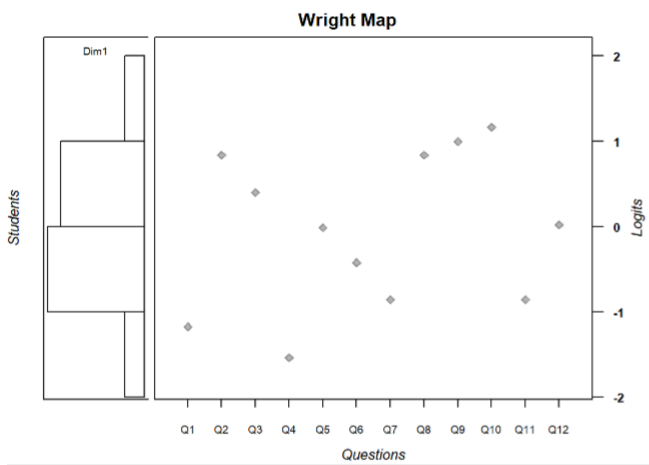

Fig. 2. Item-person map representing the location of students and questions on the same scale

and less than zero indicating a question designed for students with below-average ability. Out of all questions, the difficulty level ranged from -1.173 to 1.164 (standardized value), which indicates that our assessment was designed for students with average-level ability. Figure 1 provides question characteristics curves based on the students' ability and probability, indicating that across 12 questions, the expected curves were well aligned. Figure 2 represents the item maps using WrightMap, depicting students' locations and questions on the same scale.

Furthermore, we used infit MSQ, outfit MSQ, infit t-test, and outfit t-test to evaluate each question in the assessment. DeMars [29] suggested that items with infit MSQ or outfit MSQ values greater than 1.50 (or < 0.60) are problematic, indicating random responses by high or low performers or items not well designed for average students. As summarized in Table II, in our data, no item was flagged by the infit MSQ or outfit MSQ, indicating that the conceptual assessment is designed for students with average ability. The results also indicate that questions 1,4,6,7, and 11 were designed for students with below-average ability; questions 5 and 12 were for students with average ability, and questions 2,3,8,9 and 10 for students with above average ability, indicating a good balance of questions from medium to moderate difficulty for all students (Figs. 1, 2).

TABLE II
ITEM FIT STATISTICS FOR CONCEPTUAL ASSESSMENT FOR SEDIMENTOLOGY

Q	Total Count	Diff.	Model S.E.	INFIT		OUTFIT	
				MSQ	t	MSQ	t
Q1	34	-1.173	0.410	0.999	0.034	0.934	-0.342
Q2	34	0.840	0.393	1.104	0.634	1.181	1.032
Q3	34	0.400	0.375	1.089	0.666	1.160	1.129
Q4	34	-1.537	0.444	1.047	0.250	1.226	0.828
Q5	34	-0.012	0.368	1.116	0.948	1.131	1.051
Q6	34	-0.422	0.373	1.182	1.379	1.216	1.589
Q7	34	-0.855	0.389	1.018	0.137	0.940	-0.379
Q8	34	0.839	0.393	1.005	0.057	0.999	-0.010
Q9	34	0.997	0.403	0.831	-0.947	0.809	-10.118
Q10	34	1.164	0.415	0.771	-1.189	0.691	-1.691
Q11	34	-0.855	0.389	0.897	-0.643	0.818	-1.207
Q12	17	0.021	0.534	0.830	-0.886	0.785	-1.163

V. CONCLUSIONS AND LIMITATIONS

This work-in-progress describes the creation and validation process of the conceptual assessment test for geoscience courses to evaluate learning outcomes related to sedimentology. Using the assessment triangle approach, the paper used various reliability measures and Item response theory for two purposes: 1) to explain the process, and 2) to provide the measures that indicate the conceptual assessment reliability and show that it is designed for students with average abilities.

The results of this study must be viewed in the light of a few limitations. First, this study is part of the larger study and only presents a validation process of a conceptual assessment. Future studies can use this validated instrument for students' conceptual learning measures. Second, although the data were collected from different geoscience courses, we combined the data points for the analysis power in this study for validation purposes. Future studies can reconsider validating the conceptual assessment with a larger sample size. Third, this study didn't account for variations between courses and instructors. However, future studies may consider validating the assessment for various courses and studying its effectiveness in various settings. Fourth, this study focused on explaining the construction and validation process of conceptual assessment and didn't include students' perspectives or other process measures for evaluation. Future studies can consider using multiple data modalities for considering such needs. Fifth, for our research purposes, we limited the questions in the conceptual assessment to concepts related to sedimentology. Future studies can expand the question pool by following the same process of establishing validity and reliability.

The results of this study are novel and address the research gap by providing a validated conceptual assessment for sedimentology. Although the literature indicates existing geoscience conceptual inventories, this is the first conceptual assessment focusing on sedimentology concepts. We believe this conceptual assessment will serve the needs of many researchers and instructors for future sedimentology courses.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grant ”HSI Implementation and Evaluation Project: SedimentSketch, teaching tool in and beyond the sedimentology classroom to provide equitable and inclusive learning for Hispanic students” #2318386

REFERENCES

[1] R. G. Wolff and P. C. Benedict, “Sedimentology: general introduction and definitions: fluvial sediment and channel morphology,” US Geological Survey, Water Resources Division, Tech. Rep., 1964, https://doi.org/10.3133/70095523.

[2] A. Stepanova, S. Anwar, J. C. Laya, C. A. A. Zarikian, N. E. Martinez, and T. A. Hammond, “Board 379: Sedimentsketch, teaching tool for undergraduate sedimentology to provide equitable and inclusive learning for hispanic students,” in *2024 ASEE Annual Conference & Exposition*, 2024.

[3] D. L. Davis, “Combining active learning approaches for improving computing course outcomes at minority-majority institutions,” in *2017 ASEE Annual Conference & Exposition*, 2017.

[4] R. Cottrell, “Student outcomes in online and face-to-face classes at a hispanic-serving institution (hsi),” Ph.D. dissertation, University of Denver, 2020.

[5] D. A. McConnell, D. N. Steer, and K. D. Owens, “Assessment and active learning strategies for introductory geology courses,” *Journal of Geoscience Education*, vol. 51, no. 2, pp. 205–216, 2003.

[6] S. Sandrone, G. Scott, W. J. Anderson, and K. Musunuru, “Active learning-based stem education for in-person and online learning,” *Cell*, vol. 184, no. 6, pp. 1409–1414, 2021.

[7] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, “Active learning increases student performance in science, engineering, and mathematics,” *Proceedings of the national academy of sciences*, vol. 111, no. 23, pp. 8410–8415, 2014.

[8] R. F. Yuretich, S. A. Khan, R. M. Leckie, and J. J. Clement, “Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course,” *Journal of Geoscience Education*, vol. 49, no. 2, pp. 111–119, 2001.

[9] J. Libarkin, “Concept inventories in higher education science,” in *BOSE Conf*, 2008, pp. 1–10.

[10] J. C. Libarkin, S. W. Anderson, D. Deeds, and B. Callen, “Development of the geoscience concept inventory,” in *Proceedings of the National STEM Assessment Conference, Washington DC*, 2006, pp. 148–158.

[11] A. Jolley, F. Jones, and S. Harris, “Measuring student knowledge of landscapes and their formation timespans,” *Journal of Geoscience Education*, vol. 61, no. 2, pp. 240–251, 2013.

[12] E. D. Scribner and S. E. Harris, “The mineralogy concept inventory: A statistically validated assessment to measure learning gains in undergraduate mineralogy courses,” *Journal of Geoscience Education*, vol. 68, no. 3, pp. 186–198, 2020.

[13] J. A. N. Rhajiak, “Understanding geological time: A proposed assessment mechanism for beginner and advanced geology students at the university of british columbia,(vancouver),” 2009.

[14] J. C. Libarkin, A. U. Gold, S. E. Harris, K. S. McNeal, and R. P. Bowles, “A new, valid measure of climate change understanding: Associations with risk perception,” *Climatic change*, vol. 150, pp. 403–416, 2018.

[15] R. Glaser, N. Chudowsky, and J. W. Pellegrino, *Knowing what students know: The science and design of educational assessment*. National Academies Press, 2001.

[16] K. A. Douglas and Ş. Purzer, “Validity: Meaning and relevancy in assessment for engineering education research,” pp. 108–118, 2015.

[17] https://www.ibm.com/products/spss-statistics, 2023.

[18] D. G. Altman, *Practical statistics for medical research*. Chapman and Hall/CRC, 1990.

[19] J. Cohen, “A co-efficient of agreement for norminal scales,” *Educ Psychol Meas*, vol. 20, pp. 111–116, 1960.

[20] ——, “Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.” *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.

[21] J. R. Landis and G. G. Koch, “An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers,” *Biometrics*, pp. 363–374, 1977.

[22] S. E. Embretson and S. P. Reise, *Item response theory*. Psychology Press, 2013.

[23] S. Embretson, “Explanatory item response theory models: Impact on validity and test development?” in *Quantitative Psychology: 83rd Annual Meeting of the Psychometric Society, New York, NY 2018*. Springer, 2019, pp. 1–11.

[24] A. Robitzsch, T. Kiefer, and M. Wu, “Tam: Test analysis modules,” *R package version*, 2021, https://CRAN.R-project.org/package=TAM.

[25] D. T. Irribarra, R. Freund, and M. D. T. Irribarra, “Package ‘wrightmap’,” 2022.

[26] H. Wickham *et al.*, “Tidyverse: Easily install and load the ‘tidyverse’,” *R package version*, vol. 1, no. 1, 2017.

[27] M. Tavakol and R. Dennick, “Making sense of cronbach’s alpha,” *International journal of medical education*, vol. 2, p. 53, 2011.

[28] K. S. Taber, “The use of cronbach’s alpha when developing and reporting research instruments in science education,” *Research in science education*, vol. 48, pp. 1273–1296, 2018.

[29] C. DeMars, *Item response theory*. Oxford University Press, 2010.