Seattle
Making Value for Society
122nd ASEE Annual
Conference & Exposition
June 14 - 17, 2015
Seattle, WA

Paper ID #12198

Determining Reliability of Scores from an Energy Literacy Rubric

Dr. Chad M Gotch, Washington State University

Chad Gotch's interests center on maximizing effective and proper use of educational and psychological measurements. To this end, he studies assessment/measurement literacy among teachers, score reporting, and building validity arguments from both technical and non-technical evidence. These complimentary lines of research inform the life cycle of assessment, from development to use and policy.

Quinn Langfitt, Washington State University

Quinn is a PhD student in the School of Civil and Environmental Engineering at Washington State University. His research is mostly focused on sustainability, including work on life cycle assessment and energy literacy assessment.

Dr. Brian F French, Washington State University

Brian F. French is a Professor of Educational Psychology with an emphasis in Psychometrics and Research Methods. He is the Director of the Learning and Performance Research Center at Washington State University.

Dr. Liv Haselbach P.E., Washington State University

Determining Reliability of Scores from an Energy Literacy Rubric

Resource depletion, global climate variability, and social issues are among the challenges currently driving a need for development, assessment, and implementation of more alternative energy sources and for more efficient energy use[1]. Increased energy literacy among the general population could foster these changes. Providing factual information about costs and emissions of different energy sources can influence persons' support for different sources[2]. Additionally, access to real-time energy use information (e.g., via a smart meter) has generally resulted in decreases in energy use[3]. Therefore, energy literacy holds a place of prominence within engineering education in order to foster the ability to weigh the complex issues surrounding various energy generation sources and the capability to develop strategies for reduced energy consumption. In recognition of this prominence, the United States Department of Energy (DOE) has advocated for promotion of energy literacy through energy education in strategic plans, other documents, and various events[4,5,6]. The DOE has devoted significant efforts to the development of a guide for general energy literacy principles to serve as the basis for educational efforts[7].

Energy literacy has been measured by testing broad energy knowledge through tests and questionnaires. Such efforts have shown generally low levels of energy literacy both in children [8,9,10,11] and adults[12,13,14]. Therefore, there is a need to develop educational activities to improve energy literacy. These activities have included high school energy competitions, development of interdisciplinary curricula, and field experiences and internships. As many of these educational endeavors culminate in some type of deliverable or other artifact, an opportunity exists to supplement measurement of energy literacy via tests of knowledge with measurement through observation of project artifacts. This type of approach could then be used to examine what factors might be contributing to higher levels of energy literacy, allowing refinement of the educational activities. The development of a rubric for the evaluation of energy literacy is in progress to capture the deliverables or outcomes of a competition or course[15,16,17]. Besides providing a direct form of assessment, rubric-based evaluation of artifacts may save time and effort associated with soliciting participation and completing a test or questionnaire. The rubric approach also allows for the evaluation of past works, which creates flexibility in the timing of observations and the ability to re-assess based on a different conception of energy literacy.

Assessment of student learning is a central function of both formal and informal educational activities and settings, and has been instrumental in the advancement of engineering education[18]. Rubrics are a valuable way to assess competencies, such as those associated with energy literacy, because they allow for consistent assignment of scores based on established criteria and descriptions of performance[19]. A rubric may be used to measure competencies as demonstrated in a performance, portfolio, poster, or other educational artifacts. The production of such artifacts allows the learner to demonstrate complex skills that can be difficult to assess with multiple-choice items or concept inventories. Moreover, the learner or groups of learners may produce such artifacts in more real-world settings (e.g., internship, competition) that look different than traditional classroom assessments. While rubrics provide a means to rate performances and products across multiple dimensions and levels of proficiency, a concern in their use is that the rubric ratings are dependent upon the individuals supplying the ratings[20]. Error is introduced into the assessment process through this inherent characteristic of use. A rubric that can demonstrate a high level of consistency across raters—suggesting a small amount of measurement error—is highly valuable, as it can provide dependable assessment of energy literacy. As such, the rubric would collect quality information to support inferences about students and the effectiveness of energy literacy learning opportunities.

All uses of measurement instruments should be supported by a validation argument[21]. Such an argument links philosophical and technical considerations in support of a claim about the appropriateness of an intended use of the scores from the instrument[22]. One such consideration, which may serve as evidence for a use claim, is the reliability of scores that may be obtained from applying a measurement instrument such as a rubric. Reliability refers to the consistency in observed assessment scores[23]. The purpose of this paper is to estimate the reliability of scores from a rubric developed to assess levels of energy literacy demonstrated in projects at a high school energy competition. The present study represents a necessary effort to document performance of the rubric for refinement and to advance the study of energy literacy in many formal and informal learning environments.

Method

Instrument

A rubric-based approach for assessing energy literacy was first developed by Langfitt, Haselbach, & Hougham[15] based on the DOE's[7] framework for energy education. The structure of the rubric was borrowed from a rubric used to assess senior design projects in a civil and environmental engineering program, and then refined based on work conducted in the area of scientific understanding and writing[24]. Subsequent examinations[17] and refinements[16] of the rubric led to the rubric under examination in the present study. The rubric adopted an analytic style[19], with scores assigned to each of six energy literacy dimensions—1) Issue, 2) Solution, 3) Impacts, 4) Stakeholders, 5) Technical Concepts, and 6) Outside Information. Appendix A displays the performance descriptors associated with each score for each rubric dimension. Within each dimension, raters could give a score of 0, 1, 3, or 5. These score options reflect the conceptual distinctions between performance descriptions. For example, the additional energy literacy demonstration required to move from a score of 1 to a score of 3 was seen as greater than the additional demonstration required to move from a score of 0 to a score of 1. Each sub-principle within the DOE framework was mapped directly to one or two rubric dimensions to facilitate consistent scoring based on tangible indicators of energy literacy.

Data collection

Project posters and abstracts from an annual high school energy science/design competition held in the Pacific Northwest of the United States were used as the artifacts for scoring. Projects were completed in teams of 3-5 students with a mentor, typically a teacher at the students' school, guiding the project. Students worked in teams over the course of the school year leading up to the event. These projects were entered into one of four challenges—Behavior, Biofuels, Design, and Technology. Each participating team was required to write and submit an abstract of 50-200 words one month prior to the competition and present a poster at the competition. Abstracts were submitted electronically, while poster content was captured via photographs taken during the competition. In total, based on the 2014 competition, 183 abstracts and 134 posters were scored using the energy literacy rubric. There were 49 teams who registered for the competition and submitted abstracts, but either did not attend the event or did not present a poster.

Following the competition, three raters scored every abstract and two of those raters scored every poster. The two raters who scored every deliverable were a Ph.D. student in civil and environmental engineering and an upper level undergraduate in civil and environmental engineering. The rater who scored only the abstracts was a faculty member with a Ph.D. in education and research program in sustainability and natural resource conservation. There were no missing data as all artifacts present at the event were assessed and scored.

Prior to scoring, the three raters completed a calibration activity by individually scoring twenty abstracts from a previous competition and then discussing differences in scoring over a phone conference. Finally, four of the abstracts were rescored and discussed again. Time constraints prevented the calibration exercise from extending to the review of sample posters. This calibration activity ensured the raters were trained systematically and consistently to reduce rater error.

Analysis

Generalizability (G) theory[25] was used to determine the consistency of scores from the energy literacy rubric. Whereas Classical Test Theory, which inspired consistency estimates such as Cronbach's coefficient alpha[26,27], conceptualizes a score on a test or on a rubric such as the one employed in this study as the aggregate of a *true score* and some amount of *measurement error*, G theory allows for the decomposition of independent sources of error variance[28]. The sources, dubbed *facets* in G theory parlance, are defined by the measurement context. In the present study, project abstracts and posters served as the objects of measurement. The facets of interest were rubric dimension (*d*) and rater (*r*). The analyses adopted a *fully crossed* design, allowing for the examination of seven variance components. Each rater assigned a score along each dimension for every available abstract and poster. Using statistical algorithms based in the analysis of variance (ANOVA) model and a peer-reviewed, publicly available SAS macro[29], variance estimates were obtained for the main effects of artifacts (*a*, i.e., abstract or poster), rubric dimension (*d*), and rater (*r*). Two-way interactions of artifact-by-dimension (*a* x *d*), artifact-by-rater (*a* x *r*), dimension-by-rater (*d* x *r*), and the confounded three-way interaction (*a* x *d* x *r*) were estimated as well. The raters in this design were modeled as a random effect. That is, they were assumed to represent the typical rater that would use the rubric in the given environment with the same calibration training. This design decision is important to note as it relates to the generalizability of the findings.

The proportions of variance accounted for by each component were compared to identify relative contributions to rubric scores. Across these variance estimates, higher values signaled a stronger influence over rubric scores. A consistent, dependable rubric should be associated with a high proportion of variance accounted for by the artifact (*a*). This component represents the amount to which differences in scores reflect error-free assessments of the general energy literacy demonstrated in the artifacts. Moderate proportions of variance associated with *d* or *a* x *d* could also be present in a dependable rubric. These proportions would simply signal that artifacts varied in the extent to which energy literacy was demonstrated within the specific dimensions of the rubric. Variance associated with *r, a* x *r*, *d* x *r*, or *a* x *d* x *r* would reveal the extent to which rater inconsistencies contributed to artifact scores. As an overall measure of reliability, a generalizability coefficient, *G*, was obtained for abstract ratings and poster ratings. This estimate, like other measures of reliability, has a range of 0.00 to 1.00, with a value of 1.00 corresponding to perfect consistency in scores for making relative decisions about energy literacy demonstration in the project artifacts[28]. The relative decisions could reflect the ranking of the projects. A *G* estimate of 0.80 could be considered sufficient for support for the rubric in research, while a *G* estimate above 0.95 would be desirable in cases when the rubric is used to rank artifacts for an award[30].

Results

The variance components estimated for ratings of project abstracts are displayed in Table 1. For these ratings, proportions of variance accounted for by individual main effects or two-way interactions were all 11% or less. Collectively, these effects accounted for less than half of the variance observed in ratings. The majority of variance (53%) was contained in the three-way interaction of artifact, rubric dimension, and rater, suggesting a substantial amount of unidentifiable or random error in the scores assigned to project abstracts. This result was corroborated by a generalizability coefficient of *G*=.56

Table 1
Variance components for project abstract ratings

Source	*df*	*SS*	*MS*	*Variance estimate*	*Proportion (%)*
a	181	321.1	1.8	0.06	10
d	5	86.8	17.4	0.02	3
r	2	71.8	35.9	0.02	4
a x *d*	905	424.8	0.5	0.06	11
a x *r*	362	214.1	0.6	0.05	10
d x *r*	10	91.0	9.1	0.05	9
a x *d* x *r*	1810	515.5	0.3	0.29	53

The variance components estimated for ratings of project posters are displayed in Table 2. While the component associated with the largest proportion of variance was the three-way interaction, the amount (39%) was less than was observed with abstract ratings. Accordingly, the generalizability coefficient for poster ratings was higher at *G*=.69. With posters, a larger proportion of variance was found to be associated with true differences between artifact scores. There was very little variability attributable to differences in the amount of energy literacy observed across dimensions (*d*) or to raters seeing different amounts of energy literacy in the posters (*r*).

Table 2
Variance components for poster ratings

Source	*df*	*SS*	*MS*	*Variance estimate*	*Proportion (%)*
a	132	1418.1	10.7	0.62	26
d	5	32.4	6.5	0.00*	0
r	1	0.3	0.3	0.00*	0
a x *d*	660	1032.6	1.6	0.32	14
a x *r*	132	349.4	2.6	0.29	12
d x *r*	5	151.3	30.3	0.22	9
a x *d* x *r*	660	606.1	0.9	0.92	39

*Negative estimates were set to zero. Many methods exist for handling such estimates, but in practice are likely to produce the same result[31] (p. 85).

Discussion

This study estimated the reliability of scores from a rubric developed to assess levels of energy literacy demonstrated in projects at a high school energy competition. The analyses represent a necessary effort to document performance of the rubric for refinement and to advance the study of energy literacy. Framed within a validity argument, an application of the rubric as was conducted in this study may support only rough categorizations of energy literacy across a sample of artifacts. Scores for both abstracts and posters demonstrated a substantial amount of unexplainable variance, though results were somewhat better for judging energy literacy in project posters. Abstract content was likely too sparse to allow raters to render consistent judgments. Thus, at this time, it is suggested that artifacts that are more in-depth be used in such an assessment framework.

One way to improve the reliability of rubric scores may be to provide additional training to raters. Main effects for rater or effects for raters in combination with either artifacts or dimensions contributed to 23% of score variance for abstracts and 21% of the score variance for posters. These outcomes suggest each rater impacted an abstract's or poster's score to a non-negligible degree. The obtained variance proportions are not surprising, given differences in rater backgrounds and time constraints on rater training and calibration in the present study. Though these limitations are noted, they also reflect anticipated conditions of real-world assessment settings. That is, in such events, resources for scoring are likely to be limited. If calibration occurs on-site at the event, there may be little time for an extended session given the condensed nature of such events. If calibration occurs either before or after the event, raters may experience difficulty coordinating schedules or accommodating calibration work within demanding workloads. Calibration activities that occur remotely will likely face challenges inherent to long-distance communication (e.g., lack of face-to-face interaction). In further practical application of the rubric, to the greatest extent possible, careful attention and sufficient resources should be given to training to attempt to reduce rater effects. Such training could include ample time to discuss the DOE energy literacy framework and the expectations of each rating (i.e., 0, 1, 3, and 5) within each rubric dimension. An opportunity to iteratively score sample artifacts and discuss points of divergence should prove beneficial[32].

An additional way rubric score reliability could be improved is through the introduction of additional raters. As a general rule, adding well-trained raters should always improve score reliability. There are diminishing returns on these additions, however. A next step in the line of research to investigate and support this energy literacy rubric will be to conduct a *decision study*[25] in which reliability estimates will be extrapolated from the current data. Such a study will help determine if adequate reliability could be obtained through the addition of, for example, 1 to 3 more raters.

As interest in incorporation of sustainability concepts into engineering education is high and continues to increase[33], novel ways to engage students in the associated concepts are being developed[34]. Those approaches will need to be evaluated with a range of reliable assessment techniques, such as the rubric approach for assessing energy literacy examined in this paper. Improvements to the energy literacy rubric application process could serve to improve reliability. With that could come additional applications of the approach, such as uses for course content refinement and cases for support of particular extra-curricular activities shown to foster energy literacy, in turn improving the experience and effectiveness of energy education endeavors.

Acknowledgements

The authors gratefully acknowledge the Northwest Advanced Renewables Alliance supported by the Agriculture and Food Research Initiative Competitive Grant no. 2011-68005-30416 from the USDA National Institute of Food and Agriculture, the Center for Environmentally Sustainable Transportation in Cold Climates (CESTiCC), and Ecoworks Foundation for funding part of this work.

Bibliography

1. Black, B., & Flarend, R. (2010). *Alternative Energy*. Santa Barbara, CA: Greenwood Press.
2. Hobman, E. V., & Ashworth, P. (2013). Public support for energy sources and related technologies: The impact of simple information provision. *Energy Policy*, 63, 862–869.
3. Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*(3), 273–291.
4. United States Department of Energy. (2011). *Strategic Plan*. Retrieved from http://energy.gov/sites/prod/files/2011_DOE_Strategic_Plan_.pdf
5. United States Department of Energy. (2014). *Energy Literacy Town Hall*. Webinar retrieved from http://energy.gov/eere/education/downloads/webcast-national-energy-literacy-virtual-town-hall
6. United States Department of Energy. (2014, April 4). We're Data Jammin': Building Interactive Educational Materials to Teach Energy. Retrieved from http://energy.gov/eere/articles/were-data-jammin-building-interactive-educational-materials-teach-energy
7. United States Department of Energy. (2013). *Energy Literacy: Essential Principles and Fundamental Concepts for Energy Education*. Retrieved from http://www1.eere.energy.gov/education/pdfs/energy_literacy_2.0_low_res.pdf
8. Barrow, L., & Morrisey, T. (1989). Energy literacy of ninth-grade students: A comparison between Maine and New Brunswick. *Journal of Environmental Education, 20*(2), 22–25.
9. Gambro, J., & Switzky, H. (1999). Variables associated with American high school students' knowledge of environmental issues related to energy and pollution. *Journal of Environmental Education, 30*(2), 15–22.

10. DeWaters, J. E., & Powers, S. E. (2011). Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior. *Energy Policy, 39*(3), 1699–1710.
11. Bodzin, A. (2012). Investigating urban eighth-grade students' knowledge of energy resources. *International Journal of Science. Education, 34*(8), 1255–1275.
12. National Environmental Education & Training Foundation . (2002). Americans' low "Energy IQ:" A risk to our energy future. Retrieved from http://www.neefusa.org/pdf/roper/Roper2002.pdf
13. Bittle, S., Rochkind, J., & Ott, A. (2009). *The energy learning curve*. Retrieved from http://www.publicagenda.org/media/the-energy-learning-curve
14. Southwell, B. G., Murphy, J. J., DeWaters, J. E., & LeBaron, P. A. (2012). *Americans' perceived and actual understanding of energy*. (RTI Press peer-reviewed publication No. RR-0018-1208). Research Triangle Park, NC: RTI Press. Retrieved from http://www.rti.org/rtipress
15. Langfitt, Q., Haselbach, L., & Hougham, R.J. (2014). Artifact-based energy literacy assessment utilizing rubric scoring. *Journal of Professional Issues in Engineering Education and Practice*. Retrieved from http://dx.doi.org/10.1061/(ASCE)EI.1943-5541.0000210
16. Langfitt, Q., & Haselbach, L. (2014). *Imagine Tomorrow high school energy competition 2014: Energy literacy and biofuels literacy assessment of abstracts and posters*. Pullman, WA: Washington State University.
17. Langfitt, Q., Haselbach, L., & Hougham, R.J. (2015). Refinement of an energy literacy rubric for artifact assessment and application to the Imagine Tomorrow high school energy competition. *Journal of Sustainability Education, 8.* Retrieved from http://www.jsedimensions.org/wordpress/content/2015/01/
18. Olds, B. M., Moskal, B. M., & Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. *Journal of Engineering Education, 94*(1), 13-25.
19. McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (6th ed.). Upper Saddle River, NJ: Pearson.
20. Lane, S., & Stone, C. A. (2006). Performance assessment. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education/Praeger.
21. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
22. Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.
23. Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Educational Measurement: Issues and Practice, 10*(1), 37-45.
24. Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509–547.
25. Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice, 11*(4), 27-34.
26. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
27. Haertel, E. H. (2006). Reliability. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education/Praeger.
28. Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
29. Mushquash, C. & O'Connor, B. P. (2006). SPSS, SAS, and MATLAB programs for generalizability theory analyses. *Behavior Research Methods, 38*, 542-547.
30. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
31. Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
32. Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370-395.
33. Bilec, M.M., Hendrickson, C., Landis, A.E., & Matthews, H.S. (2011). Updating the Benchmark Sustainable Engineering Education Report: Trends from 2005 to 2010. *Proceedings of the ASEE Annual Conference and Exposition*, Vancouver, BC.
34. Johnston, L.F. (Ed.) (2013). *Higher Education for Sustainability*. Routledge, New York, NY.

Appendix A

Energy Literacy Rubric[16]

	Points			
Topic	**0**	**1**	**3**	**5**
Issue	Not addressed	Identify the issue	Frame the issue	Professionally frame the issue
Solution	Not addressed	Identify solution to the issue	Discuss a solution	Develop appropriate solution
Impacts	Not addressed	Identify broader Impacts	Discuss broader impacts	Examine broader impacts
Stakeholders	Not addressed	Identify stakeholders	Consider stakeholder perspectives	Understand and address stakeholder perspectives
Technical Concepts	Not addressed	Identify technical concepts	Discuss technical concepts	Examine technical concepts as they relate to the project
Outside Information	Not addressed	Identify basic info from outside sources or that this information exists	Discuss information from outside sources	Examine information as it relates to the project