Workshop 1: Methods for evaluating standard setting workshops

Paraskevi (Voula) Kanistra & Charalambos (Harry) Kollias

Several standard setting methods aim to quantify the "minimally adequate level of performance" (Kane, 1994, p. 425) learners need to demonstrate to successfully pass an examination, obtain a degree, or be awarded licensure to practice a trade. Such quantification of the minimal level of adequate performance is referred to as a 'cut score', a predefined performance standard that defines the level of knowledge, skills, and abilities needed to be deemed proficient enough for some area and purpose (Kane, 1994, 2001). Given the critical role of standard setting, several standard setting practitioners and researchers in the field have suggested well-established and accepted frameworks to evaluate all phases of a standard setting workshop in terms of procedural, internal, and external validity evidence (Cizek & Earnest, 2016; Hambleton, Pitoniak, & Copella, 2012; Hambleton & Pitoniak, 2006; Kane 1994). Cizek and Earnest updated evaluation framework[1] (see Cizek and Earnest, 2016) can be used to inform training materials, data collection, and data analyses of most standard setting workshops.

This workshop will utilise Cizek and Earnest's framework as a building block for evaluating standard setting workshops and will move beyond it to incorporate data analyses methods.

Intended learning outcomes

Participants will become familiar with all aspects entailed in evaluating a standard setting workshop.

Upon completion of this workshop, participants should be able to:

- apply a theoretical standard setting validation framework;
- critically appraise the methods for evaluating a standard setting workshop;
- understand the data to be collected and types of analyses needed for validation;
- conduct basic internal validity analysis using freely available software programmes;
- understand the intricacies involved in conducting a virtual standard setting workshop.

Workshop methods and contents

The workshop will engage participants through a mixture of lectures, software demonstrations, group and plenary discussions, and hands-on activities. Applying principles of collaborative learning, participants will engage in an activity employing a jigsaw strategy, whereby the standard setting evaluation framework will be broken down into smaller parts and groups of participants will be asked to adapt it to suit the standard setting method assigned and report back their adaptations. Two step-by-step demonstrations of how to conduct "interparticipant" and "decision accuracy" analyses will be performed. Participants will have the opportunity to run these analyses on their own laptops.

---

[1] The authors adapted it from Standard Setting Methods for Complex Licensure Examinations, M. Pitoniak, 2003 (Unpublished doctoral dissertation, University of Massachusetts, Amherst.

The workshop is organised as follows:

Day 1: During Day 1 participants will be provided with a brief overview of some commonly used test-centred and examinee-centred standard setting methods and the types of data collected through these methods. During the second part of the day, participants will be introduced to a comprehensive framework for evaluating standard setting workshops.

Day 2: During Day 2, the introduction to the comprehensive framework will be continued and expansion of the framework will be exemplified using the Item Descriptor (ID) Matching method. Participants will then be assigned in groups and asked to expand the framework taking into consideration one of the standard setting methods and type of data collected, previously covered in Day 1. As each group will be assigned a different method, all groups will then report back their adaptations. Towards the end of the day, participants will be introduced to free software to conduct interparticipant consistency. A step-by-step demonstration will be provided so that participants will be able to replicate on their own laptops the same analysis.

Day 3: During Day 3, participants will be introduced to another free software to conduct decision accuracy and consistency. Similar to the previous software demonstration, participants will be able to conduct the same analysis on their own laptops. The workshop will end with a discussion on best practices for conducting virtual standard setting workshops.

The agenda may be adjusted to suit the participants' expertise and needs.

Background knowledge/requirements

This workshop is intended for language assessment professionals with a special interest in standard setting either because they are planning to conduct one or because they would like to become adept at evaluating one. General knowledge of standard setting and familiarity with some common standard setting methods is assumed, but participants are not expected to have facilitated a standard setting workshop themselves. Participants new to standard setting are advised to read Zieky and Perie's primer (see references below).

Participants should bring their own laptops with pre-installed free software programmes that will be shared with registered participants prior to the workshop.

Workshop leaders

Dr. Paraskevi (Voula) Kanistra is Associate Director – Senior Researcher at Trinity College London. Voula has extensive experience conducting (virtual) standard setting workshops, aligning test instruments to the Common European Framework of Reference (CEFR). Voula's PhD thesis focused on the Item Descriptor (ID) Matching method and its application using CEFR descriptors. Additionally, she operationalised and expanded on an existing standard setting evaluation framework to include CEFR descriptors, Rasch measurements, and enhance external validation.

Dr Charalambos (Harry) Kollias is Research Director - Psychometrician in the Centre for Statistics (CfS) at the National Foundation for Educational Research (NFER) and has extensive experience in conducting and evaluating (virtual) standard setting workshops, aligning assessment instruments to international frameworks such as the European Qualifications Framework (EQF), the Common European Framework of Reference for Languages (CEFR), and the Global Proficiency Framework (GPF). In 2023, he authored "Virtual Standard Setting: Setting Cut Scores" (Peter Lang).

Both workshop leaders have presented standard setting research at local and international conferences, including the National Council of Measurement in Education (NCME), European Association for Language Testing and Assessment (EALTA), the Common European Framework of Reference (CEFR) Special Interest Group (SIG) EALTA, and the UK Rasch Users Group Meeting.

References

Cizek, G. J., & Earnest, D. S. (2016). Setting performance standards on tests. In S. Lane, m. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (2 ed., pp. 212- 237). New York and London: Routledge.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 433-470). Westport: Praeger Publishers.

Hambleton, R. K., Pitoniak, M. J., & Copella, J. M. (2012). Essential steps in setting performance standards. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (pp. 47-76). New York: Routledge.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425-461. doi:10.3102/00346543064003425.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53-88). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Zieky, M., & Perie, M (2006). A primer on setting cut scores on tests of educational achievement. ETS. Retrieved from https://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf