



14th EALTA conference

1-3 June 2017, Sèvres, France

Pre-Conference Workshops



Standard Setting for Language Tests: Methods, Issues, and Hands-on Practical Experience (workshop 2)

Workshop Leader: Steve Ferrara (Measured Progress, USA)

The international language testing community has indicated interest in considering a range of standard setting methods. For example, a committee to set standards for a new English reading comprehension test (see Figueras, Kaftandjieva, & Takala, 2013) considered the Bookmark Standard Setting Process (Lewis, Mitzel, Mercado, & Schulz, 2012) and Item-Descriptor (ID) Matching method (Ferrara & Lewis, 2012) as well as Modified Angoff. In addition, the English Language Proficiency Assessment in the US (ELPA 21; <http://www.elpa21.org/>) recently used the Bookmark method and considered ID Matching. Language testing programs often set standards using the Modified Angoff method, where standard setters estimate the percentage of borderline examinees they expect to respond correctly to each item. This method requires a probability judgment. Extensive research indicates that people make inaccurate probability judgments (e.g., Ferrara & Lewis, 2012). Holding a hypothetical borderline examinee in mind further exacerbates the cognitive load in Angoff standard setting (Ferrara & Lewis, 2012; Pellegrino, Jones, & Mitchell, 1999). This research suggests that language testing specialists might want to consider other standard setting methods for their programs. This proposed workshop is most closely aligned with 14th annual conference theme, Use and interpretation of assessment results and touches partially on the membership requested workshop topic, 1. Integrated testing and assessment: design, development, analysis, reporting of results and relevant uses.

In this workshop, participants will study and practice applying two standard setting methods that are in use for operational testing programs in the US and other countries and review two other widely used methods that are particularly relevant to assessments of writing and speaking.

Contents

Workshop participants will :

Review and discuss several standard setting methods and consider their advantages and drawbacks for language tests

Practice applying the cognitive-judgmental task required in several standard setting methods, using the same set of test items

Evaluate several standard setting methods for possible use for language tests, following a framework in Ferrara and Lewis (2012, Table 13.2): shared understanding of the knowledge and skills explicated in PLDs, shared understanding of item response demands, shared understanding of borderline examinees, criterion for ordering items in an ordered item booklet, and the cognitive-judgmental task required to recommend cut scores

Evaluate the standard setting methods for potential use for selected international language tests, to be selected ahead of the conference by workshop participants

Review and discuss special topics and considerations in standard setting

The intended learning outcomes for participants are to develop a broad understanding of standard setting method options, a deep understanding of the cognitive and judgmental requirements of several standard setting methods, and appreciation for the benefits and drawbacks of these methods for potential use in operational language testing programs. They also will learn about special issues and considerations in designing standard setting projects.

Methods

The workshop is designed around presenting concepts and demonstrating procedures, supported by printed and displayed examples; drawing out participant discussion on the content and procedures to share and enhance their understanding; hands-on, guided practice in applying the standard setting methods; and commentary to enhance understanding by the two workshop discussants.

The Workshop is designed for up to 30 participants. Previous experience in setting performance standards for language tests is desired, but not required or necessary to participate and benefit from participating in the workshop. Workshop participants will be encouraged to read a paper on each of the standard setting methods to be presented and applied in the workshop. The workshop leader will provide PDFs of the papers prior to May 29. Participants also will be encouraged to bring in their own test items, in case they would like to experiment with applying various standard setting methods with a familiar test.

Pre-workshop activities

Ferrara, S., & Lewis, D. (2012). The Item-Descriptor (ID) Matching method. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 255-282). New York: Routledge.

Figueras, N., Kaftandjieva, F., & Takala, S. (2013). Relating a reading comprehension test to the CEFR levels: A case of standard setting in practice with focus on judges and items. *The Canadian Modern Language Review*, 69(4), 359-385.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225-253). New York: Routledge.

Pellegrino, J. W., Jones, L. R., Mitchell, K. J. (1999). *Grading the Nation's Report Card: Evaluating NAEP and transforming the assessment of educational Progress*. Washington, DC: National Research Council. See <https://www.nap.edu/download/6296>

Workshop leaders

Steve Ferrara, Ph.D. is Senior Advisor for Measurement Solutions at Measured Progress, New Hampshire, USA. Prior to that, he held leadership and research positions at Pearson, CTB/McGraw-Hill, and American Institutes for Research. Steve conducts psychometric research, designs summative and formative assessments for K-12 educational achievement and English language proficiency, publishes in technical journals, and trains teachers on classroom assessment practices. Steve designed, developed, implemented, and conducted validation studies for the speaking components of the NAEP Foreign Language Assessment in Spanish and the English Development Assessment, plus prototype versions of automated speaking proficiency assessments and learning systems. He has designed and led standard settings for various achievement and language proficiency testing programs in the US and Brazil using the Modified Angoff, Body of Work, Bookmark, Item-Descriptor (ID) Matching, and Reasoned Judgment methods. Steve developed the ID Matching method with colleagues and has written and presented extensively on standard setting methods. Dr. Ferrara is on the editorial advisory boards for the US journals *Educational Assessment* and *Applied Measurement in Education* and edited *Educational Measurement: Issues and Practice*. He is co-recipient of two different research awards from the American Educational Research Association.

Neus Figueras holds a PhD in language testing from the University of Barcelona. She is catedràtica d'escuela oficial d'idiomes and has worked in the Departament d'Ensenyament de la Generalitat de Catalunya coordinating the certificate exams for the EOI. She lectures part-time at the University of Barcelona and the Universitat Pompeu Fabra. She has been

involved in a number of international research and development projects (Speakeasy, Dialang, Ceftrain) and collaborates regularly with the Council of Europe in the dissemination of the Common European Framework of Reference in relation with testing and assessment. She has published articles in the field of language teaching and assessment and is one of the authors of the Manual for Relating examinations to the CEFR (Council of Europe, 2009). She has recently published, with Fuensanta Puig, *Pautas para la evaluación del español como lengua extranjera* (2013). Edinumen. She has been a teacher trainer for over 20 years, and has given courses and presented in universities in Spain and in different European countries, in Asia and the USA. She was the first President (2004-7) of EALTA (European Association for Language Testing and Assessment), and she is now an expert member (www.ealta.eu.org).