

Установление стандарта проходного балла: валидность, надежность, практичность

Шевченко Светлана Эдуардовна¹ Прохорова Елена Фёдоровна²

¹Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия

²Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия

Standard Setting: validity, reliability, practicality

Shevchenko Svetlana¹ Prokhorova Elena²

¹Saint Petersburg State University (SPSU), St.Petersburg, Russia

²Saint Petersburg State University (SPSU), St.Petersburg, Russia

Аннотация

Анализируются факторы, обеспечивающие валидность, надежность и практичность установления стандарта, на основе результатов международной экспертизы по соотнесению Теста второго сертификационного уровня Санкт-Петербургского государственного университета (английский язык) с Общеввропейской шкалой иноязычной коммуникативной компетенции, полученных на этапе установления стандарта проходного балла для рецептивных видов речевой деятельности.

Abstract

The article focuses on the analysis of the factors ensuring validity, reliability and practicality of the standard setting, basing upon the results of the international

project on linking St.Petersburg University's Second Certificate Test (English Language) to the CEFR, that were obtained during the stage of standard setting for receptive skills.

Ключевые слова: проходной балл, соотнесение теста с уровнями Общеввропейской шкалы иноязычной коммуникативной компетенции, методы установления стандарта проходного балла, IRT, валидность и надежность

Keywords: a cut-score, linking a test to the CEFR, standard setting methods, IRT, validity and reliability

1. Процедура установления стандарта проходного балла(далее – установление стандарта) для трех разделов Теста второго сертификационного уровня



СПбГУ(далее – Тест), а именно, «Аудирование», «Чтение» и «Практическое использование языкового материала», проводилась в СПбГУ в рамках международной экспертизы по установлению соответствия Теста требованиям уровня B2 Общеввропейской шкалы иноязычной

компетенции.[CouncilofEurope, 2001]. Тем самым решалась задача достижения валидности, надежности и практичности экзамена.

2. По определению, установление стандарта требуетналичие участников и конкретного метода, который позволяет им применить свои оценки таким образом, чтобы перевести политические позиции уполномоченных организаций в определенные позиции на балльной шкале[Cizek, Bunch, с.18]. В СПбГУ была выработана позиция относительно целевого уровня владения английским языком - решение о необходимости выпускникам этого университета иметь уровень иноязычной коммуникативной компетенции не ниже B2.

3. Из десяти методов, рекомендованных Советом Европы [CouncilofEurope, 2009], была выбрана Голландская версия метода закладок (ACitoVariationontheBookmarkMethod). Это передовой метод, опирающийся на современную теорию тестов IRT. В его основе лежит однопараметрическая логистическая модель (OPLM), которая объединяет математические достоинства однопараметрической модели Раша с характеристиками двухпараметрической логистической модели, и оценивает не только параметры трудности заданий, но и учитывает коэффициенты дискриминации заданий как заданные константы [Verhelst, Glas, Verstralen].

Базовым понятием этой модели является понятие способности тестируемого, выраженной в виде латентной переменной, которая отражает соответствие между наблюдаемыми данными ответов тестируемых и теоретическим предположением вероятности правильного ответа.

6. Как показал опыт экспертизы Теста в СПбГУ, использовать Голландскую версию могут только те экзаменационные системы, которые готовы к требуемому уровню проведения процедуры установления стандарта.

Центр лингводидактического тестирования СПбГУ работал над созданием независимой экзаменационной системы шесть лет, и выбор данного метода для установления стандарта в объективных видах речевой деятельности в рамках международной экспертизы Теста неслучаен.

7. Большая часть методов при анализе тестовых заданий использует дихотомическую модель ответов экспертов на вопрос: может или не может пороговый тестируемый правильно выполнить тестовое задание. Такое упрощение задачи может вести к погрешности определения проходного балла [Reckase]. Отличительной особенностью Голландской версии метода закладок является использование понятия различных уровней вероятности правильного ответа на тестовое задание. Эксперты использовали шкалу вероятности правильного ответа (от 0,50 до 0,80), которая соотносится с 3 уровнями компетенции порогового тестируемого: (высокий, средний, и низкий (недостаточный)).

8. Эксперты принимали решение, исходя не только из своего видения тестового вопроса, то есть,

субъективной оценки, но и учитывая статистические характеристики (дифференцирующую способность и трудность) каждого тестового вопроса, основанные на реальных результатах выполнения этого вопроса тестируемыми.

9. Наличие нескольких уровней вероятности, применение объективных квалитетических данных, помогающих эксперту принять решение, позволило минимизировать ошибку оценок экспертов, что, несомненно, позволяет говорить о большой валидности, надежности и достоверности стандарта проходного балла.

10. Другим преимуществом Голландской версии метода закладок является возможность переноса стандарта проходного балла, установленного в рамках экспертизы, на другие версии Теста, что говорит о его практичности. Данный метод дает возможность проанализировать ответы тестируемых в различных вариантах Теста на единой шкале латентной переменной способности тестируемых.

11. В соответствии с требованиями процедуры Совета Европы (Council of Europe, 2009) по каждому из трех разделов экзамена проводились раунды установления стандарта и раунды перекрестной валидации. Результатом работы экспертов был количественный показатель латентной способности тестируемого, то есть его компетенции (минимального набора речевых умений и навыков, позволяющих отнести его к уровню B2) в определенном виде речевой деятельности. Среднеарифметическое значение персональных оценок экспертов считалось результатом работы всей группы экспертов в том или ином раунде.

12. После первого раунда, экспертам представлялась дополнительная

информация по среднеарифметическому значению оценки, стандартному отклонению и стандартной ошибке группы экспертов в целом. Прежде чем перейти ко второму раунду эксперты имели возможность обсудить возникшие вопросы и результаты оценивания в мини-группах.

13. Основной задачей этапа перекрестной валидации было определение уровня надежности результатов оценивания в условиях изменения процедуры установления стандарта проходного балла. Этот этап позволил показать, что при смене варианта теста, Санкт-Петербургские эксперты продемонстрировали устойчивость и стабильность оценок. А другая группа экспертов (представителей четырех стран) показала, в целом, схожие результаты, используя метод Ангоффа.

14. Важным показателем внутренней валидности процедуры установления стандарта является характеристика согласованности работы экспертов, измеряемая через относительную стандартную ошибку. Значения этого показателя, а также другие статистические данные, представленные в полном тексте статьи на сайте testing.spbu.ru/, свидетельствуют о довольно высокой точности, согласованности и устойчивости оценок экспертов.

15. Применение данного метода не только позволило получить валидные и надежные результаты установления стандарта, а значит и экспертизы, но и дало возможность продолжать использовать передовую технологию определения проходного балла, которая гарантирует максимальную объективность оценивания.



Литература

- Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press, 2001. 264.
- Council of Europe. Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): a Manual. Language Policy Division, Strasbourg, 2009. 201.
- Cizek, G.J. and Bunch, M.B. Standard Setting: a guide to establishing and evaluating performance standards on tests. Thousand Oaks: Sage, 2007. 369, –P.18.
- Reckase, M. D. A Conceptual Framework for a Psychometric Theory for Standard Setting with Examples of Its Use for Evaluating the Functioning of Two Standard Setting Methods. Educational Measurement: Issues and Practice, 2006, 25(2), –P.418.
- Verhelst, N.D., Glas, C.A.W. & Verstralen, H.H.F.M. One-Parameter Logistic Model OPLM, Cito, 1995. 164, –P.1.

References