

TEACHERS' INVOLVEMENT IN
HIGH STAKES TESTING:
DESIGNING AND IMPLEMENTING AN
ENGLISH SPEAKING EXAMINATION

Daniel Xerri & Patricia Vella Briffa
University of Malta

Concerns with high stakes testing



- Affects the quality and nature of instruction and makes **teachers feel relatively powerless** (Costigan, 2002)
- **Lack of consistency or coherence** between teachers' visions of desirable education and those articulated in high-stakes examinations (Atkin, 2007)
- The difficulty to act autonomously in the face of testing pressures forces teachers to **reinvent their professional identity** so that it's in harmony with the new testing mentality (Assaf, 2008)

Teachers as examiners



- Educators must be able to engage with an assessment system so as to contribute to policy making and **prevent a centralization of power** (Gregory & Clark, 2003)
- Through involvement **teachers develop ownership** of the procedures and criteria and understand the process of assessment (Harlen, 2005)
- The teacher, not the test, is the **primary change agent** so we must bring teacher judgement to centre stage (Klenowski & Wyatt-Smith, 2012)

Developing assessment literacy



- Teachers will be better **able to respond to the demands** of the avalanche of high stakes testing (Gulek, 2003)
- Unless the **quality of teacher judgements** is addressed they will continue being excluded from high stakes testing (Brookhart, 2013)
- Hands on involvement in high stakes test development is a means of **developing assessment literacy** (Runté, 1998; Black et al., 2011)

Context



- MATSEC Examinations Board
- Advanced English Examination
- 600 candidates (prospective university students)
- New syllabus in 2010 with new speaking component
- First sitting in May 2013
- “a measure of the candidates’ ability to speak and converse in English” (MATSEC, 2010)
- Teachers’ initiative
- Five teachers at a postsecondary school

PART 1



- Guided examiner-to-candidate conversation
- About 3 minutes
- 4 marks
- An informal interview intended as a conversation starter
- The examiner asks questions about topics such as *Work, Study, Leisure, and Career Plans*

Part 2



- Guided examiner-to-candidate conversation
- About 4 minutes
- 6 marks
- Consists of two stages:
 - a) a description of the visual prompt
 - b) a two-way exchange where topics and issues related to the visual prompt are developed

Part 3



- Candidate-to-examiner long turn
- About 3 minutes
- 8 marks
- A presentation expressed as a long turn by the candidate based on a topic
- Candidate selects topic from a list of 5
- *lifestyle, music, sport, religion, relationships, international news, environment, war, education, entertainment*
- Candidate is allowed some minutes to prepare presentation before entering the examination room

Exam specifications



- The greater the detail in the specification of content, **the more valid the test** is likely to be (Hughes, 2003)
- Syllabus constraints
- Aims and content
- Structure, timing and procedures
- Rubrics and instructions

Writing items, moderation, trialling



- Specimen materials in line with specifications
- Avoid problems with test use
- Knowledge of student cohort
- Internal and external moderation
- Trialling, observation and feedback

Rating Scale



- It would be useful if the criteria employed in the assessment of language production on tasks could be **related in a principled way** to the criteria for the teaching of a skill (Weir, 2005)
- Fluency & coherence
- Pronunciation
- Vocabulary
- Grammar

Rating scale

	Part 1		
Fluency & coherence	1-2	3	4
	descriptor	descriptor	descriptor

Calibration:

- Video recorded 50 mock tests
- Samples of performance covering the entire range of the scale

Test manual



- **Accountability** via a test manual (McNamara, 2000)
- The degree of a **test taker's familiarity** with the demands of a particular test may affect the way the task is dealt with (Weir, 2005)
- **Dynamic and evolving document** rather than monolithic (Fulcher, 2003)
- Concrete **connections between theory and practice** and thus educational (Luoma, 2004)
- Dissemination

Examiner training



- Socialise raters into a **common understanding of the scale descriptors**, and train them to apply these consistently in operational speaking tests (Fulcher, 2003)
- Trainee examiners selected from current secondary/postsecondary teachers
- Test content, procedures and assessment criteria
- Rating standardisation
- FAQs
- Cyclical and iterative (Taylor & Galaczi, 2011)

Benefits



- Empowerment via ownership
- Enhanced assessment literacy
- Beliefs, attitudes and practices
- Confidence in teachers' judgement
- Equitable examinations

References

- Assaf, L.C. (2008). Professional identity of a reading teacher: Responding to high-stakes testing pressures. *Teachers and Teaching: Theory and Practice*, 14(3), 239—252.
- Atkin, J.M. (2007). Swimming upstream: Relying on teachers' summative assessments. *Measurement: Interdisciplinary Research and Perspectives*, 5(1), 54—57.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451—469.
- Brookhart, S.M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69—90.
- Costigan III, A.T. (2002). Teaching the culture of high stakes testing: Listening to new teachers. *Action in Teacher Education*, 23(4), 28—34.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Pearson Education.
- Gregory, K., & Clarke, M. (2003). High-stakes assessment in England and Singapore. *Theory Into Practice*, 42(1), 66—74.
- Gulek, C. (2003). Preparing for high-stakes testing. *Theory Into Practice*, 42(1), 42—50.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies, *Curriculum Journal*, 16(2), 207—223.

References

- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65—79.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- MATSEC Examinations Board. (2010). *Advanced Matriculation syllabus English 2013*. Retrieved from http://www.um.edu.mt/__data/assets/pdf_file/0015/108024/AM10.pdf
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Runté, R. (1998). The impact of centralized examinations on teacher professionalism. *Canadian Journal of Education*, 23(2), 166—181.
- Taylor, L., & Galaczi, E. (2011). Scoring validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp.171—233). Cambridge: Cambridge University Press.
- Weir, C.J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Contact



- daniel.xerri@um.edu.mt

- patricia.vella-briffa@um.edu.mt