BUILDING UP A POOL OF STANDARD SETTING JUDGES: PROBLEMS SOLUTIONS AND INSIGHTS

RITA GREEN & CAROL SPOETTL

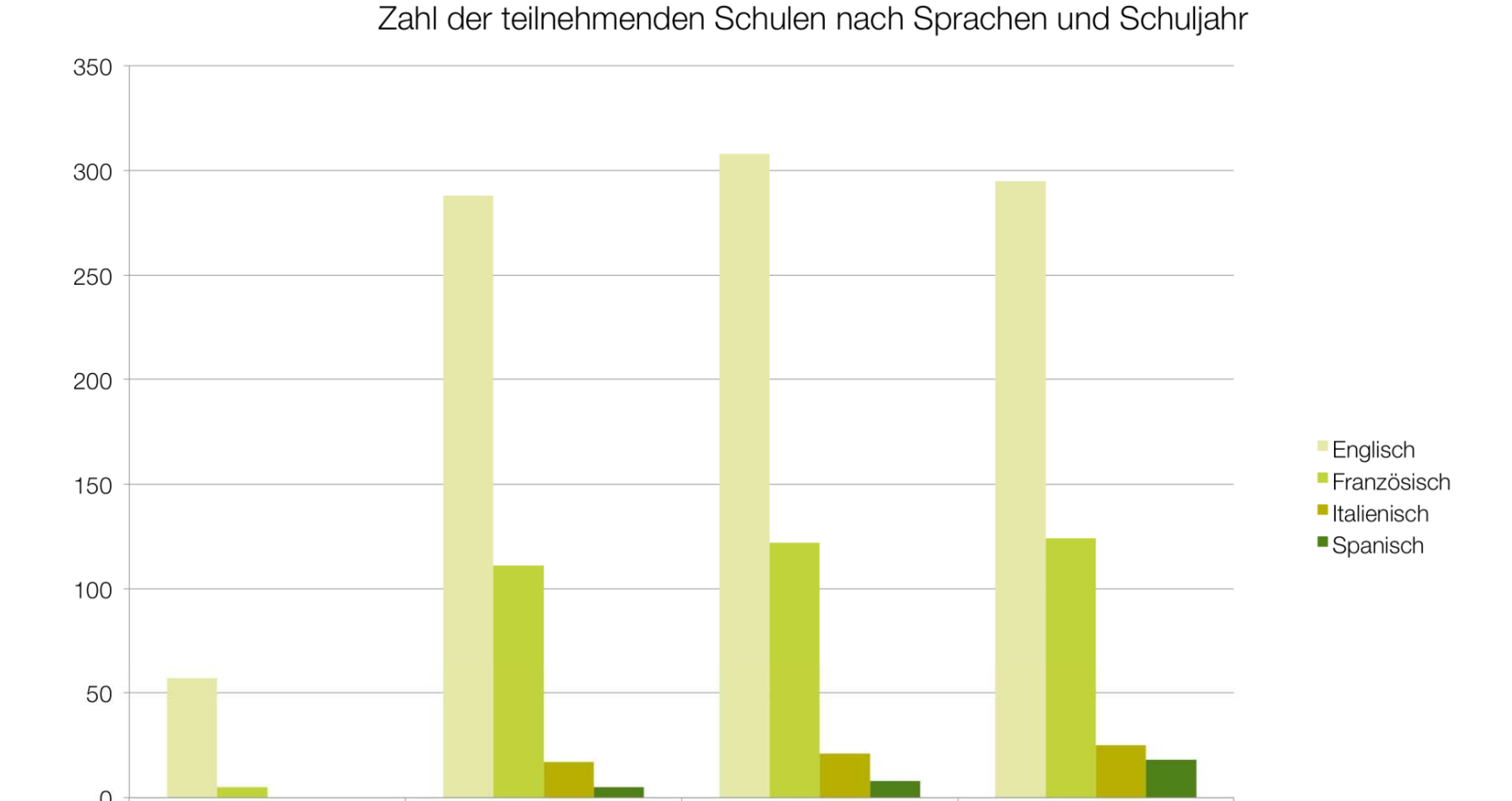
OVERVIEW

- Austrian SRP project & standard setting needs
- Procedures for setting up a pool of judges
- Problems faced
- Solutions reached
- Insights gained

THE AUSTRIAN SRP PROJECT

Standarisierte Reifeprüfung (SRP)

- Austrian parliament passes education bill October 2010
- Standardised tasks for all Austrian schools by 2013 in
  - German
  - Maths
  - Foreign languages
- Innsbruck University Ministry's partner foreign languages
- 2 CEFR levels: B2 and B1
- 4 Languages: English, French, Italian and Spanish
- Item writer training and task development began 2007
- First live administration in selected pilot schools 2008



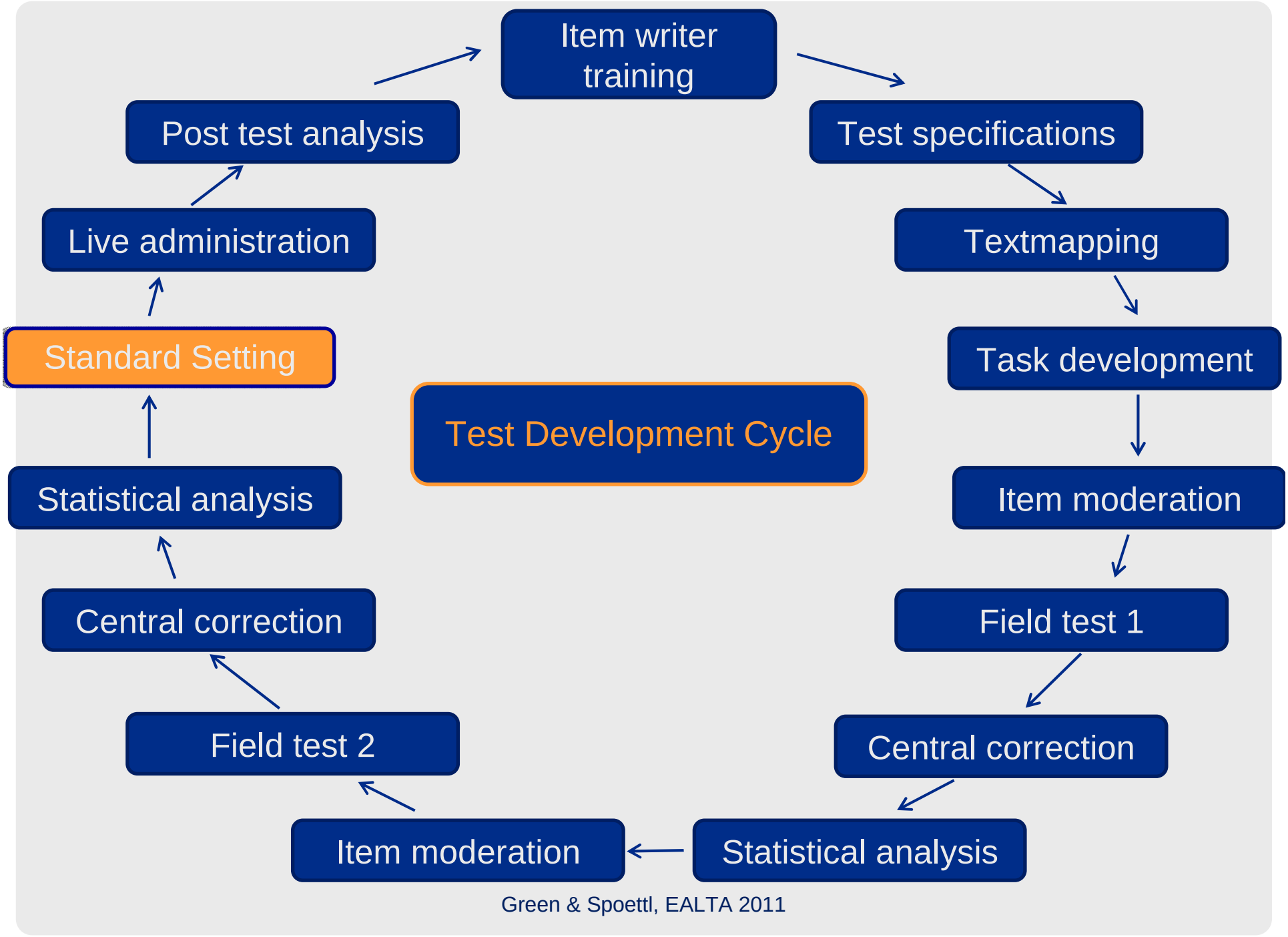
	2008	2009	2010	2011
Englisch	57	288	308	295
Französisch	5	111	122	124
Italienisch	0	17	21	25
Spanisch	0	5	8	18

Bundesinstitut bifie Wien | Zentrum für Innovation & Qualitätsentwicklung

SRP LIVE ADMINISTRATION MAY 2011
SKILLS AND CEFR LEVELS

	Reading		Language in Use		Listening	Writing [2012]
English	B2		B2		B2	B2
2nd FL	4yrs	6yrs	4yrs	6yrs		
French	B1	B2	B1	B2	B1	B1
Italian	B1	B2	B1	B2	B1	B1
Spanish	B1	B2	B1	B2	B1	B1

Green & Spoettl, EALTA 2011



Green & Spoettl, EALTA 2011

SRP STANDARD SETTING NEEDS

- December 2007
  - English Listening and Reading
  - French Listening
- December 2008
  - French Listening and Reading
  - Italian Listening
- Jan 2009: English Listening & Reading
- December 2009 & December 2010
  - English Listening & Reading;
  - French Listening & Reading;
  - Italian Listening
  - Italian Reading (Dec 2010 only)
- September 2010
  - English & French Writing

OBJECTIVES IN SETTING UP THE POOL

- raise awareness of the role of standard setting in high stakes tests
- train judges across the 4 languages to be called upon to participate in standard setting sessions
- gain further insights into the judgement making process which could be fed back into task development work
- for reasons of transparency
- encourage confidence in the product for stakeholders in their specific work situation

STANDARD SETTING JUDGES

"Who made thee a judge?" Raymond and Reid 2008

1)"training participants is a crucial step in helping to ensure that the cut scores the participants set are meaningful"

2) "…an important task in any cutscore study is choosing participants who are qualified to make the necessary judgments and who are representative of all qualified participants."

Zieky, Perie, Livingston 2008

JUDGES: THE SRP STAKEHOLDER INVOLVEMENT

- The BIFIE [www.bifie.at]
- The Austrian school inspectorate
- AHS headmasters
- AHS teachers
- Austrian universities
- Austrian teacher training experts
- External panelists [IQB, CIEP]

WHY THESE PARTICULAR STAKEHOLDER GROUPS?

- School inspectorate have to deal with:
  - headmasters who have to implement change
  - parents who have children affected by the new Matura
  - teachers who have to deliver the new Matura
- Headmasters have to deal with:
  - apprehensive parents
  - uneasy staff
  - concerned students
- Teachers:
  - familiarity with target test population
  - closer familiarity with immediate teaching context
- An international / non-Austrian judge:
  - the European perspective vs. the national perspective

BUILDING UP A POOL: PROCEDURES (1): JUDGES

Pool required:

- 4 languages:
  - English, French, Italian & Spanish

Pool size

- Aimed for 40
  - 10 participants per language
  - Representatives from each stakeholder group

ZIEKY, PERIE & LIVINGSTON'S RULES FOR TRAINING

- Participants must be qualified to decide what level of knowledge and skills measured by the test is necessary
- Participants must be representative of all qualified people
- Participants must be representative of demographic characteristics
- For educational accountability tests
  - Judges should mostly consist of school teachers, curriculum specialists and school administrators
  - Representatives from different school types
  - Representatives from different types of students e.g. gifted and special needs
  - Representatives from different regions
- Group size: 12-18 but minimum 8

PROBLEMS ENCOUNTERED (1)

- Problem 1
  - The higher up the hierarchy, the fewer the representatives available from the romance languages (F, I, S)
- Problem 2
  - The time factor (range of other commitments)
- Problem 3
  - Diversity in stakeholders' knowledge of the CEFR
- Problem 4
  - Lack of knowledge of standard setting and the role of statistics

PROBLEMS ENCOUNTERED (2)

- Problem 5
  - Awareness of the possible conflict of interests: personal – own children; professional – own students
- Problem 6
  - Lack of a common L1 across language groups
- Problem 7
  - Working with various translated versions of the CEFR
- Problem 8
  - Participants' understanding of the construct underlying the tests

SOLUTIONS (1)

- Problem 1: Representativeness – romance languages

  Some judges turned out to have more than 1 language and were willing to standard set in 2 languages

  Some judges turned out to fulfil two job categories e.g. teacher and PH consultant, teacher at school and university

- Problem 2: The time factor

  Participants were given the workshop dates 2 years in advance; off-peak times in school were identified

  On-going discussions with the Ministry to prioritise standard setting sessions over other educational commitments.

SOLUTIONS (2)

- Problem 3: Stakeholders' knowledge of the CEFR

  Familiarisation exercises were given at the beginning of every workshop

  Sample items were done individually and discussed in plenary from past live test administrations

- Problem 4: Lack of standard setting knowledge

  Input session on purpose, procedure and judges' role were provided in workshop 1

SOLUTIONS (3)

- Problem 5: Conflict of interest
  Participants were required to complete a judges' background information form

- Problem 6: Lack of common L1 across groups
  Simultaneous presentations made in English and German using 2 data projectors

SOLUTIONS (4)

- Problem 7: Four versions of the CEFR

  For training purposes we reduced the versions to English / German

- Problem 8: Understanding of the construct

  Each session began with an input on the particular skill being focused on in the workshop

GENERAL INSIGHTS

- Insight 1: The novelty of standard setting
  - Professional development & networking
- Insight 2 : Doing the test
  - Getting the right answer was a priority
- Insight 3: Rating item difficulty
  - CEFR level vs. own pupils' ability vs. judges' own ability
- Insight 4: Multiple matching tasks
  - Most challenging task type across languages
  - Mismatch between judges' perceived difficulty vs. live statistics
  - Age factor

FACTORS AFFECTING JUDGEMENT OF DIFFICULTY LEVEL (LISTENING)

- 6 listening tasks
- B2 (English); B1 (French/Italian)
- 3 test methods:
  - multiple choice (2)
  - multiple matching (2)
  - note form (2)
- Judges
  - English (n=12)
  - French (n=10 to 12)
  - Italian (n=4)

PROCEDURE

Judges were asked to what extent their judgements were affected by:

- the topic
- the test method
- the accent(s) of the speaker(s)
- the speed of delivery
- their own ability to do the task

FACTORS AFFECTING JUDGEMENT OF DIFFICULTY LEVEL

Based on all 6 tasks, the findings were as follows:

Factor	English	French
Speed of delivery	1	1
Accent(s) of speaker(s)	2	5
Test method	3	2
Topic	4	3
Own ability to do task	5	4

JUDGES' FINAL COMMENTS

- I learnt a lot about rating. I now know that simply referring to the CEF descriptors is not enough.
- Gaining more insight into features that could contribute towards predicting difficulty of items
- The difficulty of the item depends on a whole myriad of things therefore judging it as B1, B2, … might turn out quite difficult
- Knowing the descriptors is one thing – applying them - that's another story

POST-TRAINING BENEFITS

Live standard setting session

- better understanding of their role as judges
- higher quality of discussion, more focused
- increased consultation of CEFR descriptors during rating
- better understanding of the role of trial statistics to inform their judgement

CONTACT DETAILS

ritagreen_tdta@hotmail.com

carol.spoettl@uibk.ac.at