



Institut zur Qualitätsentwicklung
im Bildungswesen

EALTA 2007
Sitges

*The challenge of constructing a
valid writing assessment system
based on the CEF*



Claudia Harsch
IQB, HU Berlin
claudia.harsch@iqb.hu-berlin.de

Rita Green
UK
ritagreen_888@hotmail.com

- Background
- Training
- Rating scale
- Feasibility study
- Feedback
- Next steps

- 2004: German National Educational Standards for school leaving exams for German, Mathematics, Natural Sciences and the First Foreign Language
- Lower track, 9th graders, 14-16 years old:
CEF level A2
- Middle track, 10th graders, 15-17 years old:
CEF levels B1/B1+

A2: The learners can

- give a very short, basic description of events
- write short, simple notes (...) referring to immediate needs and necessities (...)

B1/B1+: The learners can

- write uncomplicated, detailed texts on a number of different topics on their areas of interest
- phrase simple standardised letters and e-mails, such as enquiries (...)

- Institute for Educational Progress
- Research-oriented institute at the Humboldt-University of Berlin
- Supporting all 16 federal states of Germany
- Core mandate:
 - * monitor the educational development within the German school system
 - * evaluate and further develop the National Educational Standards
 - * advance their implementation in schools

- Educational Standards to be operationalised through a series proficiency tests:
 - Reading / Listening / Writing / Speaking
- Team of language testing experts, teachers, didactic experts, and psychometricians
- Group of teachers:
 - * One from each federal state of Germany
 - * Covering all school types (lower / middle / higher track, comprehensive schools)

- Teachers in Germany receive no special training in language test development
- 9 workshops focusing on
 - * familiarisation – the Standards and CEFR
 - * principles of language testing
 - * test specification development
 - * task development
 - * trialling and data analysis
 - * task revision

- The workshops were supplemented by
 - * regular regional group meetings
 - * work on the IQB platform
- Main focus: A2 and B1/B1+ in line with the Standards
- Tasks developed to target a specific level
- Templates based on the test specifications and the Dutch Construct Group work

Writing Template: Part 1



section 1: general data			
Itemwriter:		Team:	--▼ (S1 En)
Type of template:	Writing Compr. V1	Version/date:	
prompt:			
Copyright:			
Type of prompt:	--▼	Source:	--▼
Nature of content:	--▼	Length of input:	words
Topic area:	--▼		
output:			
Discourse type:	<input type="checkbox"/> narrative <input type="checkbox"/> descriptive	<input type="checkbox"/> instructive <input type="checkbox"/> expository	<input type="checkbox"/> argumentative <input type="checkbox"/> phatic
Text type:	--▼		
Time allowed:	--▼	Length of output:	--▼
functions:			
<input type="checkbox"/> to appeal	<input type="checkbox"/> to attract interest	<input type="checkbox"/> to comment	
<input type="checkbox"/> to describe	<input type="checkbox"/> to evaluate	<input type="checkbox"/> to excuse	
<input type="checkbox"/> to explain	<input type="checkbox"/> to express experiences / feelings		
<input type="checkbox"/> to express opinions/viewpoints/preferences			
<input type="checkbox"/> to give advice/ask for advice			
<input type="checkbox"/> to give arguments	<input type="checkbox"/> to give directions	<input type="checkbox"/> to inform/to ask for information	
<input type="checkbox"/> to instruct	<input type="checkbox"/> to invite	<input type="checkbox"/> to make arrangements	
<input type="checkbox"/> to persuade/convince	<input type="checkbox"/> to provide support	<input type="checkbox"/> to state advantages/disadvantages	
<input type="checkbox"/> to summarize	<input type="checkbox"/> to tell a story	<input type="checkbox"/> to thank/apologize	

Writing Template: Part 2



comprehension level:	
A1	<input type="checkbox"/> 1 Write groups of words / simple isolated phrases and sentences, ... <input type="checkbox"/> 2 Fill in simple forms... <input type="checkbox"/> 3 Ask for or pass on personal details in written form... <input type="checkbox"/> 4 Write a short simple postcard... <input type="checkbox"/> 5 Write simple phrases and sentences about themselves ...
A2	<input type="checkbox"/> 1 Write groups of words / phrases / sentences ... <input type="checkbox"/> 2 Write short and simple (formulaic) notes, messages and posters, ... <input type="checkbox"/> 3 Write very simple personal letters and emails ... <input type="checkbox"/> 4 Write short simple texts ...
B1	<input type="checkbox"/> 1 Write straightforward detailed texts ... <input type="checkbox"/> 2 Write notes and messages <input type="checkbox"/> 3 Write personal letters, simple formal letters and emails... <input type="checkbox"/> 4 Write short reports/articles ...
B2	<input type="checkbox"/> 1 Write clear, detailed connected texts ... <input type="checkbox"/> 2 Write a description ..., a review, an essay or report ... <input type="checkbox"/> 3 Express news and views and relate to those of others. <input type="checkbox"/> 4 Write letters conveying degrees of emotion ... <input type="checkbox"/> 5 Write an essay or report ... <input type="checkbox"/> 6 Evaluate different ideas or solutions to a problem.
C1	<input type="checkbox"/> 1 Write clear, well-structured texts (reports, articles, essays, descriptions, ... <input type="checkbox"/> 2 Write clear, detailed, well-structured and developed descriptions and texts... <input type="checkbox"/> 3 Write formal and informal letters ...

- Following documents were taken into account:
 - * the test specifications
 - * the German Standards
 - * the CEFR descriptors
 - * any existing scales based on the CEFR i.e. Manual, Into Europe, DESI, Finnish
- Initial drafts of the scale were drawn up showing a synthesis of this information, colour-coding being used to illustrate the original sources

SEE1 A2 Rating Scale, v2



Grammar	<ul style="list-style-type: none"> Uses some simple structures and sentence patterns (parataxis and some more complex patterns in a learnt repertoire) correctly but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless it is usually clear what he/she is trying to say. 	GA – p114 DESI – Grammar Level 2 CEFR Manual Table 5.8
	<ul style="list-style-type: none"> Has problems with subordinate clauses. Errors regarding word order occur. Performance shows <i>noticeable</i> mother tongue influence. 	DESI – Grammar Level 2
	<ul style="list-style-type: none"> Errors may sometimes cause misunderstandings [accuracy] Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations [range] 	CEFR Manual Table 5.8
	<ul style="list-style-type: none"> Local errors that do not hinder understanding 	Into Europe, p131 = A2 / B1

Early drafts → project team. In light of their feedback:

- perceived redundancies removed
- number of criteria confirmed
- certain criteria divided into two to make them more accessible i.e. range / accuracy in grammar and vocabulary
- descriptors re-worded to remove perceived ambiguity

Scales → the team, IQB colleagues, other project experts for comment.

Further revisions made in light of received feedback.

Grammar	<p>Range</p> <ul style="list-style-type: none">+ Uses some simple structures [present/past/future, simple modals e.g. can/may/must; auxiliaries e.g. verb to be/have: [see Bildungsstandards p14] and sentence patterns e.g. questions/answers, negatives/positives, commands, suggestions <p>Accuracy</p> <ul style="list-style-type: none">+ Still systematically makes basic mistakes e.g. tends to mix up tenses, forget to mark agreement, has problems with subordinate clauses, errors in word order; nevertheless it is usually clear what he/she is trying to say.+ Local errors [e.g. single elements within a sentence, wrong verb form] which do not hinder understanding are likely to occur frequently [nearly every sentence]+ Global errors, which hinder understanding at the sentence level, are likely to occur+ Performance shows <i>noticeable</i> mother tongue influence.
----------------	---

Criteria of assessment:

- task fulfilment
- organisation
- grammar
- vocabulary

Decisions

- a uni-level approach to grading the tasks in line with specific level at which they had been developed
- performance would be graded on a fail / pass / pass plus basis

- 36 tasks ranging from A1 to C1 levels
- 3 different types of schools
- each test taker attempted 3 tasks
- administered by team across Germany
- approximately 40 scripts per task
- feedback questionnaires and trial reports

Simultaneous trial of both scales and tasks

- Initial work

Pre-selection of benchmarks for one task at each level illustrating performances at fail / pass / pass plus

- Training

- * Grading of BMs by teachers, giving justifications
- * Grading of further scripts on same tasks

In light of seminar experience, teachers

- identified further tasks which appeared to work
- where possible identified benchmarks
- identified tasks which did not work
=> set aside for revision

In regional groups, teachers

- graded remaining tasks in February
- double marking, familiar level

Rating Sheet, v2



Rating Sheet

Assessor: _____ Candidate no: _____ Task: _____ Task level: _____

					Comments / Justification
Task fulfilment	No attempt 0	Fail 1	Pass 2	Pass-Plus 3	
Organisation	No attempt 0	Fail 1	Pass 2	Pass-Plus 3	
Grammar	No attempt 0	Fail 1	Pass 2	Pass-Plus 3	
Vocabulary	No attempt 0	Fail 1	Pass 2	Pass-Plus 3	
Final level awarded:	No attempt 0	Fail 1	Pass 2	Pass-Plus 3	

Trial findings revealed need for:

- further revision of scales
- re-assessment of benchmarks
- revision of some tasks
- revision of *task fulfilment* expectations

Also revealed

- small B2/C1 population, therefore insufficient scripts
- need for extra trial: grade 12, 5 B2/C1-tasks
- May-June 2007: teachers identified benchmarks, graded scripts, provided feedback on B2-C1 scales

Aims

- Trial training / rating approach with student teachers
- Gain insight into scales and criteria
- Get feedback on accessibility of handbooks, benchmarks, software [DPC]

Procedure

- 2 tasks, A2 “Lost dog” / B1 “Keeper for a day”
- 6 raters, student teachers of English, proficient in writing in English
- 2 training sessions, rating of 50 scripts per task over two weeks

- Approach appears feasible
- Scales seem to be usable and applicable:
Agreement after the first training session, both tasks,
overall rating

No of ratings (n=275)	Agreement
26 (9.5%)	3 out of 6
53 (19.3%)	4 out of 6
74 (26.9%)	5 out of 6
122 (44.4%)	6 out of 6

- Further explore behaviour at descriptor level

- Teachers / Hamburg student raters [n=21]
- Feedback suggested reasonable satisfaction with:
 - * the degree to which the construct and standards were reflected in the scales
 - * the wording, and divisions within, the criteria of assessment
 - * the specific level approach adopted
 - * the division of performance into fail/pass/pass plus

- the number of times a script had to be read to assign a grade generally increased with the perceived difficulty level of the task
- approximately 80% found the benchmarks very useful
- over half indicated a medium level of confidence with the grades they finally awarded the tasks, while a further 42% indicated a high level of confidence

- *How could your job as a rater be made easier / more reliable when using this scale?*
 - * more benchmarks
 - * further discussions

- *How could the scale be improved?*
 - * reducing the text to simple phrases/keywords

- Rating of scripts from the pilot study:
 - * further benchmarks
 - * extended training with larger group of raters
- Compare grades of the uni-level with a multi-level approach
- Wider use of the scale in Germany



Institut zur Qualitätsentwicklung
im Bildungswesen

Telefon + 49 + (0)30 + 2093 - 5508
+ 49 + (0)821 + 598 - 5750
Telefax + 49 + (0)30 + 2093 - 5336
E-mail claudia.harsch@IQB.hu-berlin.de
Website www.IQB.hu-berlin.de

Postadresse

Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

Sitz

Jägerstraße 10- 11
10117 Berlin

***Thank you for
your attention!***

Rita Green
UK

ritagreen_888@hotmail.com

HUMBOLDT-UNIVERSITÄT ZU BERLIN

