

DEVELOPING APPROPRIATE MONOLINGUAL ASSESSMENT CRITERIA FOR TESTS OF MANY LANGUAGES: ISSUES AND CHALLENGES

Dr Jo Lewkowicz

Advisor

University of Warsaw, Language Certificate

1
Examinations

OUTLINE

- Aim of presentation
- Theory underpinning the development of the rating scales used for the Certificate Exams
- The Qualities of the rating scales used
- Issues, challenges and limitations of the rating scales used

- Focus will be on writing scales but reference will be made to oral scales where relevant.

WHAT IS A RUBRIC?

- Qualitative description of language placed on a rating scale used to assess performance
- Raters match performance with relevant description to generate a score

IMPORTANCE OF RATING SCALES

- Acknowledging McNamara (1996), Weigle (2002: 109) notes:

‘the scale that is used in assessing performance tasks ... represents, implicitly or explicitly, the theoretical basis upon which the test is founded; that is, it embodies the test (or scale) developer’s notion of what skills of abilities are being measured by the test.’
- Given that at each level performance of spoken & written language is assessed, ensuring the rubrics used embody the constructs being assessed is crucial

FUNCTIONS OF RATING SCALE

- Alderson (1991) distinguished between different functions of rating scales:
 - User-oriented
 - Assessor-oriented
 - Constructor-oriented
- Pollitt & Murray (1996) added:
 - Diagnosis-oriented
- Certificate Exams rubrics are assessor-oriented to guide assessors and facilitate scoring

EXTRACT OF B2 WRITING SCALE

Treść wypowiedzi	Organizacja/styl/spójność	Poprawność językowa	Zasób słownictwa
Wypowiedź jest w pełni zgodna z tematem; poprawnie dobrane, interesujące argumenty za i/lub przeciw, właściwie rozwinięte właściwa konkluzja	Wypowiedź jest bardzo dobrze zorganizowana, tekst powiązany w akapity z użyciem właściwych podział na rozwinięte akapity; zastosowano różnorodne środki spójności tekstowej; właściwy styl i rejestr	W pełni poprawne użycie struktur językowych odpowiednich do poziomu; poprawna ortografia	Słownictwo całkowicie odpowiednie do zadania, różnorodne, poprawne i liczne kolokacje, frazeologizmy

CHARACTERISTICS OF RATING SCALE

- Holistic: single score awarded for each piece of writing
- Multiple-trait: of focus are those aspects of writing that are specific to the writing task set and specific for the test

MATCHING TASK WITH RATING SCALE

○ Task:

- Violent television programmes and video games make people more violent in real life.

○ Sample task

- Representative of tasks for this level as described in test specifications:

„student potrafi napisać związany, spójny tekst, wyrzający jego poglądl ub opinię’

MAPPING THE RATING SCALE ON CEFR

B2	Vantage or upper intermediate	<ul style="list-style-type: none">• Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization.• Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party.• Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
----	-------------------------------	--

OVERVIEW OF B2 RATING SCALE FOR WRITING

Pkt.	Treść wypowiedzi	Organizacja/styl/spójność	Poprawność językowa	Zasób słownictwa
33-40	Wypowiedź jest w pełni zgodna z tematem; poprawnie dobrane, interesujące argumenty za i/lub przeciw, właściwie rozwinięte właściwa konkluzja	Wypowiedź jest bardzo dobrze zorganizowana, tekst powiązany w akapity z użyciem właściwych podział na rozwinięte akapity; zastosowano różnorodne środki spójności tekstowej; właściwy styl i rejestr	W pełni poprawne użycie struktur językowych odpowiednich do poziomu; poprawna ortografia	Słownictwo całkowicie odpowiednie do zadania, różnorodne, poprawne i liczne kolokacje, frazeologizmy
25-32	Wypowiedź jest zgodna z tematem; argumenty są poprawnie dobrane wywód jest właściwie przeprowadzony	Wypowiedź jest właściwie zorganizowana, akapity powiązane i rozwinięte; styl i rejestr języka są właściwe, zastosowano poprawne środki spójności tekstowej	Poprawne użycie struktur językowych; poprawna ortografia; nieliczne błędy nie zakłócają komunikacji	Słownictwo odpowiednie do zadania, w miarę różnorodne; obecność kolokacji i frazeologizmów
17-24	Wypowiedź jest na ogół zgodna z tematem i zawiera większość poprawnych argumentów, jednak są one niepełne lub nie w pełni rozwinięte (zawierają nieistotne informacje); wywód nie jest całkowicie przejrzysty dla czytelnika; limit słów nie został w pełni wykorzystany	Wypowiedź jest w miarę poprawnie zorganizowana; styl i rejestr języka zawierają drobne uchybienia; zastosowano dość ubogie środki spójności tekstowej	Główne struktury językowe poprawnie stosowane; ortografia wystarczająco poprawna, aby nie zakłócać komunikacji; występują znaczne błędy gramatyczne	Słownictwo podstawowe, ale właściwe dla zadania, nieliczne kolokacje i frazeologizmy
9-16	Wypowiedź nie jest wystarczająco zgodna z tematem, nie zawiera wymaganych elementów lub są one niepoprawne, ponadto nie są one rozwinięte lub nieistotne dla tematu; limit słów nie został wykorzystany w znacznym stopniu	Wypowiedź jest mało poprawnie zorganizowana, akapity słabo rozwinięte; styl i rejestr języka często są nieodpowiednie; nie zastosowano środków spójności tekstowej	Struktury językowe często błędnie używane lub zbyt proste; liczne błędy ortograficzne; utrudniony odbiór ze względu na ilość i charakter błędów	Słownictwo nie zawsze odpowiednie do zadania, ograniczenie do najprostszych frazeologizmów
0-8	Wypowiedź nie jest zgodna z zadaniem; brak argumentacji, opinii; limit słów nie został wykorzystany w rażącym stopniu	Wypowiedź nie jest zorganizowana lub brakuje części tekstu; nieodpowiednie do odbiorcy styl i rejestr języka, brak środków spójności tekstowej	Istotne i częste uchybienia w stosowaniu struktur językowych; liczne błędy ortograficzne blokują zrozumienie	Słownictwo nieodpowiednie, niezróżnicowane, powtórzenia, uproszczenia

DEFINING FEATURES OF CERTIFICATE EXAM RATING SCALES

- Stepwise within each level
- Levels do not expect ‘perfection/complete mastery’ – but proficiency for that level
- No pass mark
- Range of marks within each band
- Allow for assessing ‘jagged profiles’
- Encourages raters to consider all traits being assessed

EVALUATION OF THE RATING SCALES

- The qualities of good rating scale are:
 - Represent explicitly the skills it intends to measure
 - Clear and explicit
 - Useable and interpretable
- To what extent do the criteria match these qualities?

ISSUES AND CHALLENGES

- Scales are subject to individual interpretation
 - To limit this, there are regular standardization sessions but these have several limitations
- Scales articulated in Polish may not reflect fully the complexities of the specific languages tested
 - This could usefully be investigated
- Test scores are the product of the whole measurement process – prompt, raters, scoring criteria and procedure: need to recognise this
 - More needs to be done to validate the scales used
- Marking pressures may impact on the reliability of test scores awarded
 - Calibration of raters would facilitate identifying those raters who consistently score high or low / identify wayward markers

CONCLUSION

- Developing rating scales for different performance skills and different levels of language proficiency and for a large cross-section of languages, is no mean feat. It is an iterative process which needs to be under constant scrutiny if we are to ensure valid, reliable, practical and fair criteria by which we judge our students.

THANK YOU