Development of an Empirically Informed Rating Rubric in a Large-Scale Testing Context

Alex Volkov, Adam Gesicki, Jake Stone

Paragon Testing Enterprises

University of British Columbia

Agenda

✓ Descriptor pool

✓ Q-Sort

✓ Descriptor checklist

✓ Banding the descriptors

✓ Piloting

Empirically Informed Scale

Sources of data:

- ✓ Publicly available rating rubrics
- ✓ Current rating scale
- ✓ Interviews with experienced raters
- ✓ CELPIP-General language ontology

Descriptor pool iterations

✓ 790 unique descriptors

✓ 455 stand-alone positively worded descriptors with no modifiers

✓ 213 descriptors relevant to CELPIP general context

Q-Sort

✓ Four language experts

✓ Five conceptual categories

✓ Message, Accuracy, Complexity, Discourse Structure, and Task

Descriptor Checklist

✓ 152 descriptors

✓ 100 samples

✓ 10 external experts

✓ Rating on the descriptor level

Descriptor Checklist

Some students have discussed going straight into full time work, but I would like to advise against this. Generally a college education can get you further in your career then attempting to enter the fulltime workplace now. College will prepare you for experiences that are directly related to your chosen field. It may seem like a lot of work, but with a little push now I know that it will pay off for you greatly in the end.

I wish you the best of luck in your future travels and studies. Please feel free to ask me any questions you may have regarding your choices ahead.

Sincerely,

███████

Sample 5Y79

Yes ⬤

	True	False
1. Justifies ideas with specifics	⬤	○
2. Gives simplified ideas	○	⬤
3. Uses vocabulary that is ambiguous	⬤	○

Descriptor Checklist Analysis

✓ Not monotonically increasing data

- Provides simplified ideas

- Uses simple sentences

- Repeats ideas

✓ Factor analysis

with results converted to IRT-like "metrics" (a-, b-parameters) for ease of understanding

Descriptor Checklist Analysis

Output

	A	B	C	F	G	I	
1	DCLCode	aParam	SE	bParam	SE2	p-val4	Descriptor
2	AOA0214	-0.214	0.172	-1.11	1.152	0.335	Makes errors in spelling which do not impede readability
3	AOA0205	-2.852	0.541	-0.212	0.104	0.042	Demonstrates lexical control of complex language
4	AOA0224	-12.781	6.743	-0.164	0.076	0.03	Readability is enhanced by writer's grammar
5	AOA0225	-5.339	1.167	-0.076	0.087	0.383	Shows grammatical control
6	AOA0226	-6.061	1.436	-0.041	0.082	0.616	Uses correct grammar
7	AOA0195	-13.556	8.48	0.071	0.07	0.308	Produces clear text
10	AOA0208	-4.22	0.865	0.075	0.089	0.4	Demonstrates control of syntax
11	AOA0204	-5.243	1.202	0.082	0.083	0.325	Uses clear language
12	AOA0207	-3.413	0.672	0.105	0.095	0.271	Uses correct collocations
13	AOA0228	-5.853	1.321	0.142	0.08	0.076	The reader finds it easy to read
14	AOA0223	-5.654	1.298	0.24	0.082	0.003	Easily read despite mistakes
15	AOA0196	-2.521	0.476	0.296	0.11	0.007	Uses appropriate vocabulary
16	AOA0206	-3.095	0.604	0.432	0.101	0	Uses accurate vocabulary
17	AOA0216	-12.216	7.172	0.474	0.08	0	Produces intelligible writing
19	AOA0201	-7.077	2.203	0.573	0.092	0	Reader understands despite the errors
20	AOA0197	-7.246	2.318	0.613	0.095	0	The reader can understand the sentences

Banding the Descriptors

Internal Panel

Pilot Scale

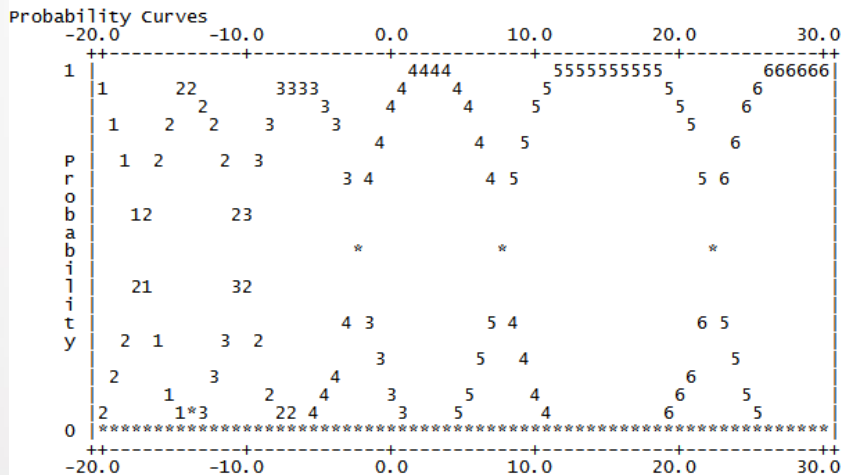| Message	Complexity	Accuracy	Discourse Structure	Task
The response is vague The writer does not state the purpose The writer repeats ideas *The message is unclear* — **1**	Uses repetitive language Uses only basic words Makes repetitive lexical choices *Uses disconnected repetitive words* — **1**	The reader cannot comprehend the sentences *The reader cannot understand the text* — **1** Grammatical mistakes predominate The reader is disrupted by the language structure — **2**	The reader has to reread the text Uses simple isolated phrases and sentences Writer's response is missing paragraphing *The structure and format are confusing* — **1**	Off-topic Task not addressed Misses all points *Does not address the given task* — **1**
Reader needs to infer Provides scarce ideas Shifts from one idea to another *Provides a simplified message* — **2**	Uses different words and expressions Uses simple sentences Uses high-frequency vocabulary *Uses simple repetitive sentences* — **2**	Makes errors in vocabulary Reader's comprehension is disrupted by vocabulary errors Sentence fragments predominate *Syntax and lexis impede understanding*	Creates a cohesive text Uses awkward format Links a series of simple elements into a sequence *The message has an awkward structure and format* — **2**	Addresses the task Misses points Mentions content points *Addresses the given task* — **2**
Provides examples and details Justifies ideas with specifics Gives reasons to support the position Supports ideas — **3**	Uses a range of vocabulary Attempts complex structures Uses a range of sentence forms Uses a range of grammatical features — **3**	Makes basic mistakes Reader gets distracted by errors Shows control of simple lexical structures Makes awkward word choices — **3**	Creates an organized text Uses appropriate format Provides a sense of closure Provides an overall structure to the text — **3**	Addresses parts of the task more effectively than others Responds to the task with development Writer's response is on topic *Completes the given task* — **3**
The reader understands the message Provides relevant ideas *Provides a message with key points and some detail*	*Uses an array of appropriate lexical and syntactic forms* Uses colloquial and natural phrases Uses complex structures	Uses awkward language structure *Syntax and lexis make comprehension challenging*	*The message has an appropriate structure and format* Transitions are clear and effective	Completes the tasks with relevant information Achieves the task *Completes and develops the given task* — **4**
Develops or elaborates ideas Organizes ideas Makes the purpose clear Uses appropriate register The reader can easily understand the main points Creates a coherent text — **4**	Uses complex sentence forms Uses vocabulary flexibly Uses low frequency words Has rich repertoire of words *Uses complex syntactic and lexical forms* — **4** Uses advanced terms	Displays instances of precise language Reader understands despite the errors Shows control of simple language structures Produces intelligible writing Uses suitable vocabulary The reader finds it easy to read *Occasional errors do not impede understanding* — **4**	Structures the response clearly Uses helpful paragraphing Uses correct punctuation *The clear format of the message facilitates reading* — **4** The discourse structure encapsulates the purpose of the Creates a sense of movement from beginning to end The key elements are appropriately located and segmented	
Provides a coherent message with clear meaning in the appropriate tone Provides ideas which flow together Achieves intended effect on the reader Fulfills her purpose Elaborates main ideas with relevant information — **5**	Uses sophisticated grammar elements Demonstrates idiomaticity *Uses sophisticated syntax, lexis with clarity and precision* — **5**	Uses correct grammar Uses correct collocations Uses precise language Uses clear language Produces clear text *The writer's language is clear and precise despite surface errors* — **5** Readability is enhanced by writer's grammar	Creates seamless transitions between points Discourse structure enhances comprehension *The structure and format of the message enhances comprehension* — **5**	
Uses appropriate tone Uses polite forms *The tone is appropriate, purpose is clear, the message achieves its intended effect*				
Provides convincing information Displays audience awareness Expresses figurative meaning Displays competent style Expresses precise meaning Clearly expresses her meaning *Expresses precise meaning with convincing information* — **6**		Consistently uses precise language Demonstrates control of syntax Shows lexical control *The writer's control of grammar and vocabulary facilitates understanding* — **6**		
Conveys shades of meaning Ideas are mature or sophisticated Examines different perspectives *Presents sophisticated ideas and distinguishes subtle shades of meaning* — **7**				

Piloting with Raters

✓ 12 experienced raters

✓ 500 samples, double rated

✓ Feedback panel and Think Aloud protocols

TESTING ENTERPRISES
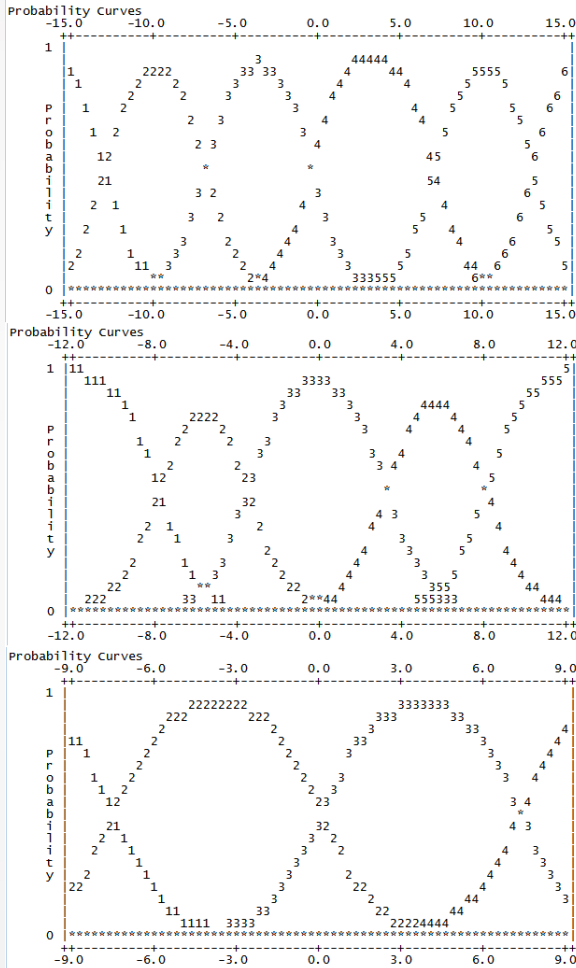Paragon

Probability Curves
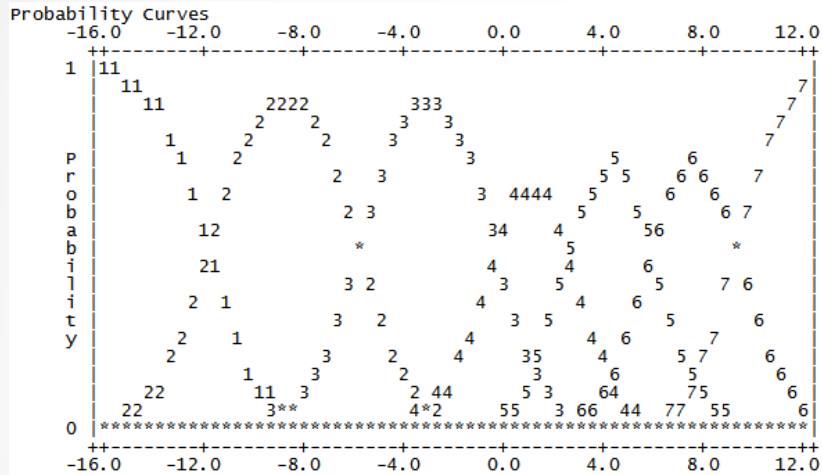
Message

Complexity

Probability Curves

Accuracy

Discourse structure

Task

Conclusion

Empirically informed rating scale

Language Ontology and Framework

Qualitative data	→	**Quantitative data**

External researchers
Expert panels
Feedback from raters

← Factor analysis
MFRM
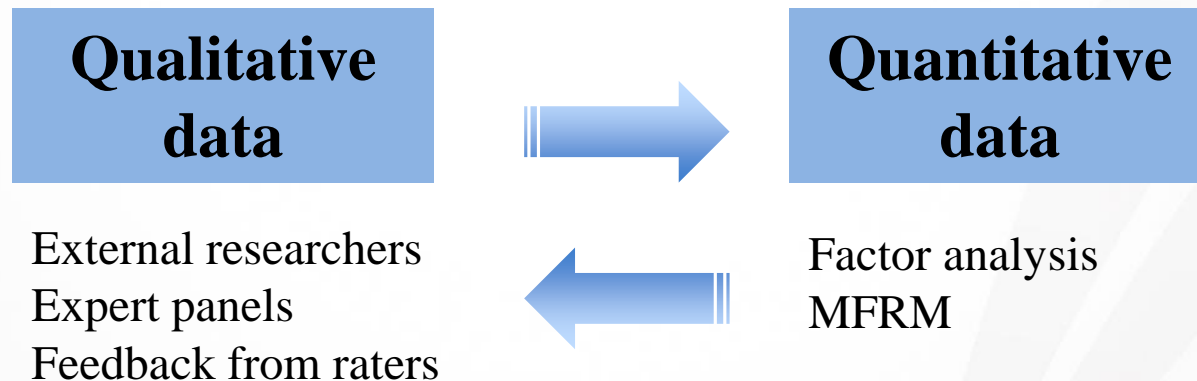
Thank you!

Development of an Empirically Informed Rating Rubric in a Large-Scale Testing Context

Questions?