

Collecting procedural evidence through comprehensive evaluation survey forms of panelists' impressions

Charalambos Kollias
Hellenic American University

 Hellenic American University

Background to the study

Communicative English Teaching Test (CETT™) Examination

- Paper 2: Language Awareness
- Paper 1: Methodology

Standard Setting Methods:

Two Variants of the Angoff Method

- Mean Item Estimation Method (polytomously scored items)
- Modified Angoff Method (multiple choice items)

Comprehensive panelist' evaluation surveys (Cizek, 2012)

11 Surveys administered

- Five: Paper 2
- Five: Paper 1
- One: Final Survey

Brief Panelists' Background: (N=14)

Current Position: Teachers (12)

Teacher Trainers (11); Administration (3)

Teaching Experience (years): 6 – 10 (3);

11 – 15 (2); 16-20 (3); Over 20 (6)

Teacher Training Experience (years):

None (2) ; 1 – 5 (6); 6 – 10 (3); 11-15 (1);

16 – 20 (2)

Training Students for CETT™:

No (5); Yes (9)

CETT™ Familiarity: Not at all (2); A little (2); Familiar (3); Very Familiar (7)

 Hellenic American University

Administration of surveys at specific junctures

1. Following Introduction/Orientation Activities
2. Following Training in Selected Method
3. Following Completion of Round One Ratings/Judgments
4. Following Round One Feedback and Completion of Round 2 Ratings/Judgments
5. Following Round Two Feedback and Completion of Round 3 Ratings/Judgments
6. Final Evaluation at Conclusion of Standard Setting Workshop

(Cizek, 2012)

Evaluation elements addressed through surveys

Explicitness (procedural validity)

- Purposes and processes clearly and explicitly articulated to participants

Implementation (procedural validity)

- Reasonable procedures conducted systematically and rigorously

Feedback (procedural validity)

- Participants confidence in standard setting process and resulting standard(s)

Reasonableness (external validity)

- Final performance standards seen as feasible or realistic

(Cizek, 2012; Sireci et. al., 2012)

Evaluation # 1: CETT™ Panelists' Feedback

KEY:	1 = Strongly Disagree; 2 = Disagree; 3 = Agree; 4 = Strongly Agree	Paper 2: (Mean)	Paper 1: (Mean)
1.	The orientation session provided a clear overview of the purpose of the standard setting of the CETT™.	3.79	3.93
2.	The orientation session answered questions I had about standard setting.	3.93	3.93
3.	I have a good understanding of my role in the standard setting activity.	3.86	3.93
4.	Reviewing the CETT™ content helped me understand the standard setting task.	3.86	3.86
5.	Experiencing the CETT™ helped me understand the difficulty, content, and other aspects of the paper.	3.86	3.93
6.	I have a good understanding of the CETT™ Borderline Performance Level Descriptors (BPLDs) .	3.36	3.93
	...		

Panelists' Feedback grouped

Understanding of how to complete ratings	Paper 2: (Mean)	Paper 1: (Mean)	Evaluation Form
I understood how to complete my Round 1 ratings.	3.86	3.93	#3 (End of Round 1)
I understood how to complete my Round 2 ratings.	3.93	4.00	#4 (End of Round 2)
I understood how to complete my overall cut score recommendations.	3.93	3.93	#5 (End of Round 3)
Confidence	Paper 2: (Mean)	Paper 1: (Mean)	Evaluation Form
I am confident in my Round 1 ratings.	3.50	3.71	#3 (End of Round 1)
I am confident in my Round 2 ratings.	3.71	3.79	#4 (End of Round 2)
I am confident in my overall cut score recommendations.	3.79	3.86	#5 (End of Round 3)

Panelists' level of reliance on information

KEY: H = Heavily; M = Moderately; S = Slightly; N = Not at all; B = Missing Data

	H	M	S	N	B	
10. My experience taking the test.	12	2	0	0		
11. My own experiences with real students.	10	3	0	1		
12. The Performance Level Descriptors (PLDs).	5	8	1	0		
13. The descriptions of the borderline groups.	7	7	0	0		
14. The item performance information (e.g., p-values, means).	7	5	1	0	1	
15. The small group discussions.	8	6	0	0		
16. The panel discussions.	14	0	0	0		
17. The normative information (i.e. ratings of other participants).	4	10	0	0		
18. The consequences information (i.e. impact data).	6	6	2	0		

(N = 14)

 Hellenic American University

Panelists' level of reliance on information cont.

Level of Reliance on Information KEY: H = Heavily, M = Moderately, S = Slightly N=Not at all (+) source relied upon most (-) = source relied upon least

	H	M	S	N	+	-
10. My experience taking the test	12	2	0	0	4	0
11. My own experiences with real students	10	3	0	1	3	1
12. The Performance Level Descriptors (PLDs)	5	8	1	0	1	2
13. The descriptions of the borderline groups	7	7	0	0	0	0
14. The item performance information (e.g., p-values, means)	7	5	1	0	2	0
15. The small group discussions	8	6	0	0	0	1
16. The panel discussions	14	0	0	0	1	1
17. The normative information (i.e. ratings of other participants)	4	10	0	0	0	1
18. The consequences information (i.e. impact data)	6	6	2	0	0	5
			Total	11	11	

 Hellenic American University

Suggestion for future adaptation of survey form

Level of Reliance on Information	Rank order the source of information you relied on the most to the least. Place one for the <u>most</u> and nine for the <u>least</u> .
10. My experience taking the test	
11. My own experiences with real students	
12. The Performance Level Descriptors (PLDs)	
13. The descriptions of the borderline groups	
14. The item performance information (e.g., p-values, means)	
15. The small group discussions	
16. The panel discussions	
17. The normative information (i.e. ratings of other participants)	
18. The consequences information (i.e. impact data)	

Suggestion for future use

Cross-reference to ... empirical check (Kane, 2001)

- Intrapanelist consistency (Hambleton et. al, 2012)

Key: 4 = Strongly Agree; 3 = Agree; 2 = Disagree ; 1 = Strongly Agree	Paper 2: (Mean)	Paper 1: (Mean)	Evaluation Form
The final group – recommended cut score for CETT™ fairly represents the minimal level of performance for CETT™ examinees	3.86	3.79 *	#6 (Final Evaluation)

* P6 found recommended cut score too low

Intrapanelist consistency: Paper 1 (P6)

Changes in ratings across rounds

	Round 1 & Round 2 Difference in recommended mean cut score	Round 2 & Round 3 Difference in recommended mean cut score
P6	0	-0.7

Spearman correlations with item-difficulty (p-values)

	P6	Group
Round 1	.52*	.79*
Round 2	.52*	.85*

* Correlations significant at $p \leq 0.01$

Investigating P6 responses

Level of Reliance on Information	KEY: H = Heavily, M = Moderately, S = Slightly N=Not at all (+) source relied upon most (-) = source relied upon least					
	H	M	S	N	+	-
10. My experience taking the test	√					
11. My own experiences with real students	√					
12. The Performance Level Descriptors (PLDs)	√				√	
13. The descriptions of the borderline groups	√					
14. The item performance information (e.g., p-values, means)		√				
15. The small group discussions	√					
16. The panel discussions	√					
17. The normative information (i.e. ratings of other participants)		√				
18. The consequences information (i.e. impact data)			√			√

References

Cizek, G. J. (2012). The forms and functions of evaluations in the standard setting process. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 165 - 178). New York: Routledge.

Hambleton, R. K., Pitoniak, M.J. & Copella (2012). Essential steps in setting performance standards in educational tests and strategies for assessing the reliability of results. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 47 - 76). New York: Routledge.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53 - 88). Mahwah, N J: Erlbaum.

Sireci, S.G., Randall, J, & Zenisky, A. (2012) Setting valid performance standards on educational tests. *CLEAR Exam Review*, 23 (2), 18 - 27.

 Hellenic American University

THANK YOU!

 Hellenic American University