



# Measuring Academic Misconduct: Evaluating the Construct Validity of the Exams and Assignments Scale

**Kenneth D. Royal\***, Keven Flammer

Department of Clinical Sciences, North Carolina State University, Raleigh, USA

**Email address:**

kdroyal2@ncsu.edu (K. D. Royal)

**To cite this article:**

Kenneth D. Royal, Keven Flammer. Measuring Academic Misconduct: Evaluating the Construct Validity of the Exams and Assignments Scale. *American Journal of Applied Psychology*. Special Issue: Psychology of University Students. Vol. 4, No. 3-1, 2015, pp. 58-64.

doi: 10.11648/j.ajap.s.2015040301.20

**Abstract:** The purpose of this study was to evaluate the psychometric properties of the Examinations and Assignments Scale (EAS), a newly designed instrument intended to capture perspectives about the severity of a variety of potential misconduct actions and behaviors, and examine evidence for construct validity. A total of 140 veterinary medical students completed the survey in the spring of 2015. Psychometric results indicate the EAS is a psychometrically-sound instrument capable of producing valid and reliable measures of misconduct severity. Substantive results and implications are also discussed.

**Keywords:** Psychometrics, Measurement, Validity, Academic Misconduct, Cheating, Veterinary Medical Education

## 1. Introduction

Virtually all students have some intuitive sense of what constitutes academic misconduct. This is especially true for actions and behaviors that are particularly egregious or severe. However, for other types of academic misconduct perceptions often vary considerably, as the notion of what constitutes academic misconduct is an inherently subjective construct. Burrus, McGoldrick and Schuhmann [1] noted that students often do not understand what misconduct is and suggests that implicit and explicit definitions of cheating tend to create a disparity between faculty and students. Further, vague conceptions of academic misconduct are not particularly helpful for college and university faculty and administrators that are expected to uphold academic integrity and fairly enforce academic conduct policies. Therefore, it is necessary that faculty and administrators are able to understand students' perceptions of misconduct, including both the nature and severity of a variety of potential offenses. This information is invaluable for creating institutional policies that clearly define academic misconduct, preserve academic integrity and enforce compliance [2]. The purpose of this study was to evaluate the psychometric properties of the Examinations and Assignments Scale (EAS), a newly designed instrument intended to capture perspectives about the severity of a variety of potential misconduct actions and behaviors, and examine validity evidence for construct

validity.

## 2. Methods

### 2.1. Participants

The EAS was administered to veterinary medical students at a large public university in the United States. A census sample approach was employed with 376 student across all four years of the Doctor of Veterinary Medicine (DVM) program invited to take part in the survey. Ultimately, 140 students participated yielding a response rate of 37.2%. Table 1 provides a demographic breakdown of DVM sample participants.

*Table 1. Student participants by DVM program year.*

DVM Year	N	%
Class of 2015	27	19.3
Class of 2016	42	30.0
Class of 2017	43	30.7
Class of 2018	28	20.0
Gender		
Male	28	20.0%
Female	112	80.0%

In an effort to investigate non-response bias, chi-squared goodness-of-fit tests were performed. Results indicate the sample was not significantly different from the larger student population based on gender,  $\chi^2(1, N = 140) = 0.01, p = .920$ , or

by program year,  $\chi^2(3, N = 140) = 3.11, p = .375$ .

## 2.2. Instrumentation and Measure

The EAS consists of 23 items intended to measure the extent to which a variety of actions and behaviors constitute misconduct in the context of course examinations and assignments. A 7-point semantic differential scale (1 = Not Misconduct to 7 = Severe Misconduct) was used to capture participant responses. A 7-point scale was used because measurement precision was a foremost concern. The EAS consists of a combination of previously used items [3-5], as well as some new items. More specifically, authors of the aforementioned studies modified items to fit their specific instructional environments and modified the rating scales to fit their unique study purposes. Modifications made to the items used in this study were also modified on the basis of instructional environment and relevance. Several new items were included to focus on clinical aspects of a veterinary curriculum.

Data were collected using Qualtrics online survey software program. The survey window to respond was open for approximately one month. Non-responders were emailed up to two additional reminders in an effort to improve participation rates. Institutional Review Board (IRB) approval was granted to conduct this study.

## 2.3. Rasch Measurement

Many measurement experts consider Rasch models to be the “gold standard” approach for mental measurements as they are the only family of measurement models that possess the properties necessary for objective measurement [6-9]. Advantages of Rasch models and how they overcome the limitations of traditional statistical models has been well-documented in the psychometric literature. Readers unfamiliar with Rasch models are encouraged to see [6, 10-11] for further information about Rasch models, particularly in the context of survey research.

In brief, Rasch models are psychometric models that assert a person with a greater amount of a latent trait will always have a higher probability of endorsing any given item than a person with a lesser amount of the latent trait. Likewise, a more difficult item to endorse will always have a lower probability of endorsement than a less difficult item. Rasch models are unidimensional measurement models that measure a primary latent trait relative to the difficulty of a task or assignment. In the case of survey research, the primary latent trait measured for persons typically is an individual’s tendency to endorse a given item. Item difficulty estimates are also produced based on each item’s tendency to be endorsed. Person and item measures are then placed along a common linear continuum so the relationship between each can be better understood. Because Rasch models are probabilistic models, one’s likelihood of endorsing an item becomes a logistic function of the relative distance between the person and item measures.

We elected to use the Rasch Rating Scale Model (RRSM)

[12] for the analysis of survey data as it is well-suited for polytomous data. According to the RRSM, the probability of a person  $n$  responding in category  $x$  to item  $i$ , is given by:

$$P_{.ni} = \frac{\exp \sum_{j=0}^x [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^m \exp \sum_{j=0}^k [\beta_n - (\delta_i + \tau_j)]}, x = 0, 1, \dots, m$$

where  $\tau_0 = 0$  so that  $\exp \sum_{j=0}^0 [\beta_n - (\delta_i + \tau_j)] = 1$ ,  $\beta_n$  is the person’s position on the variable,  $\delta_i$  is the scale value (difficulty to endorse) estimated for each item  $i$  and  $\tau_1, \tau_2, \dots, \tau_m$  are the  $m$  response thresholds estimated for the  $m + 1$  rating categories.

Winsteps [13] measurement software estimated the parameters for the model using joint maximum likelihood estimation (JMLE) procedures [14].

## 3. Results

### 3.1. Traditional Statistical Results

Traditional statistical results are presented in Table 2.

### 3.2. Psychometric Properties of the EAS

#### 3.2.1. Dimensionality

Dimensionality was investigated by performing a Rasch-based principal components analysis (PCA) of standardized residual correlations. In total, 64.4% of the Rasch dimension was explained. The largest secondary dimension explained 4.7% of the variance and had an Eigenvalue of 3.0, indicating a magnitude of about 3 items. The variance explained by the items was 37.3%. This is approximately nine times the variance explained by the secondary dimension. Evidence suggests the EAS was primarily unidimensional, and sufficiently unidimensional for a Rasch measurement analysis.

#### 3.2.2. Reliability

Rasch models provide two estimates for reliability, real and model. “Real” reliability is the worst case reliability estimate, and “model” reliability is the best case estimate. Reliability estimates for the EAS were .92-.93, indicating highly reproducible measures. Separation refers to the number of statistically distinguishable levels within the data. The separation value for this study was 3.72, indicating about four levels of statistically distinguishable levels are discernible.

#### 3.2.3. Rating Scale Effectiveness

The rating scale was evaluated to determine its adequacy and functioning. Results indicate respondents made full use of the scale, although fewer respondents used the lower extreme side of the scale. The infit and outfit mean square values for categories 1 and 2 were a bit high, but this is not necessarily problematic as most items were expected to receive higher ratings. Linacre [15] notes structure

calibrations should advance in a stepwise manner. Here, structure calibrations advance as anticipated. Finally, a visual inspection of probability curves was performed to discern if any categories were behaving in unexpected ways (see Figure

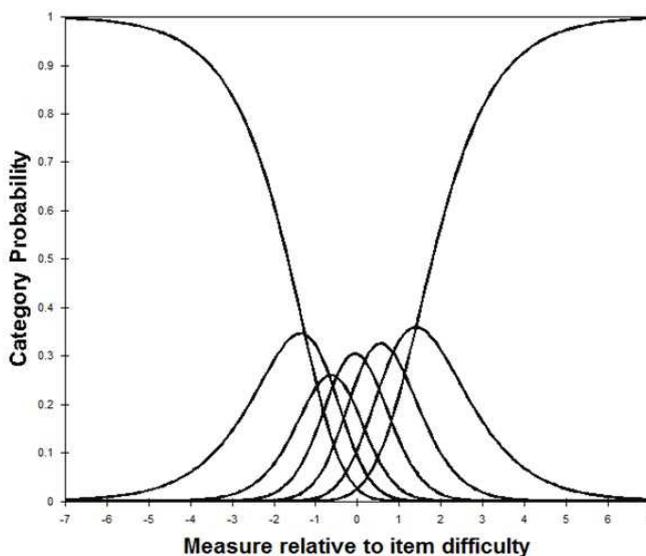
1). Results indicate the probability curves are sufficiently separated from one another, suggesting that respondents were indeed able to appropriately distinguish seven distinct levels of ratings.

*Table 2. Results of Traditional Statistical Analysis.*

Item	Mean	SD
1. Copying from another student during a quiz or examination	6.59	0.94
2. Using unauthorized cheat sheets or other materials during a quiz or examination	6.61	0.94
3. Using direct quotes from other sources without giving proper reference	5.61	1.41
4. Asking another student for the questions and/or answers to an examination which he/she had taken and you will take in the future	6.06	1.17
5. Using unauthorized test questions from a previous year, including materials found on public websites.	4.97	1.55
6. Posting unauthorized information about exams, assignments, quizzes, etc. on social media	5.66	1.48
7. Listing unread sources in the bibliography of an assignment	4.41	1.68
8. "Making up" sources for bibliographic citation	5.71	1.50
9. Working with another student on a quiz or homework assignment that was assigned as individual work.	5.12	1.67
10. Taking a graded quiz or examination for another student	6.79	0.73
11. Providing information about an exam that was intended to be confidential.	5.99	1.28
12. Missing class or lab due to a false excuse	4.31	1.64
13. Claiming to have attended class when you actually did not	4.86	1.72
14. Using a false excuse to postpone an exam	5.65	1.47
15. Removing items from a reserved reading file so that others will not have an opportunity to review them	6.13	1.27
16. Changing a response after a paper/exam/quiz was graded, then reporting that there had been a misgrade and requesting credit for your altered response	6.71	0.79
17. Permitting another student to look at your answer sheet during a quiz or examination.	6.44	1.03
18. Claiming to have handed in a paper/examination when in reality you did not	6.44	0.94
19. Failing to prepare adequately for a group assignment or laboratory	3.25	1.63
20. Doing less than your fair share in a group project or a laboratory	3.73	1.66
21. Visiting a professor after an exam or at the end of the semester to bias his/her grading	4.54	2.01
22. Presenting your clinical skills book for signing without actually completing the skill	5.05	1.70
23. Listing false completions on your online clinical skills completion summary	5.35	1.60

*Table 3. Rating Scale Diagnostics.*

Rating Category	n	%	Infit MnSq	Outfit MnSq	Structure Calibration	Category Measure
Not Misconduct	92	3	1.51	1.78	NONE	-2.68
(2)	142	5	1.31	1.63	-1.26	-1.36
(3)	189	6	0.94	0.92	-0.68	-0.63
(4)	328	11	0.95	0.91	-0.50	-0.06
(5)	475	15	0.91	0.77	0.18	0.56
(6)	648	21	0.92	0.68	0.86	1.41
(7) Severe Misconduct	1202	39	1.16	1.09	1.40	2.81



*Figure 1. Probability Curve of Rating Scale Categories.*

### 3.2.4. Item Measure Quality

Item measure quality was investigated by a number of indicators and procedures. Item difficulty measures exhibited considerable variation (Range = -2.24 to 1.95 logits) and the standard error for each item measure was quite small ( $M = .10$ ,  $SD = .04$ ). This indicates a good spread of item difficulty estimates that are relatively stable. Infit and outfit mean square statistics indicate the extent to which each item fit the model's expectation. Wright and Linacre [16] suggested reasonable values should range between .60 and 1.40 to ensure noise-free calibrations. Here, no items fell below this suggested range, but four items exhibit somewhat inflated fit statistics. According to Wright and Linacre, items that "overfit" are typically overly predictable, but do not degrade measurement quality. Items exceeding values of 2.0, however, do introduce noise that may distort measurements. Typically, outfit statistics are more concerning than infit statistics because these estimates reflect values at the extreme ends of the response vector [11]. Given these guidelines, it

appears that only item #21 may potentially exhibit noise and should be regarded as a candidate for removal. Point-measure correlations indicate the extent to which the items

discriminate. All estimates are positive, suggesting the items discriminate well. A full breakdown of item indicators are presented in table 4.

Table 4. Item Quality Indicators.

Item	Difficulty Measure	Error	Infit Mnsq	Outfit Mnsq	Pt. Measure Correlation
Q1	-1.40	.15	1.14	0.66	.54
Q2	-1.49	.15	1.41	0.86	.47
Q3	0.10	.09	1.12	1.28	.55
Q4	-.38	.10	0.79	0.75	.67
Q5	0.70	.08	0.69	0.74	.75
Q6	0.06	.09	0.87	0.75	.71
Q7	1.12	.07	0.94	0.93	.72
Q8	0.00	.09	1.20	1.05	.62
Q9	0.58	.08	1.29	1.62	.61
Q10	-2.24	.21	1.32	0.70	.40
Q11	-.30	.10	0.84	0.72	.68
Q12	1.22	.07	0.68	0.65	.78
Q13	0.79	.08	1.15	1.18	.69
Q14	0.09	.09	0.95	0.82	.71
Q15	-0.55	.10	1.81	1.56	.37
Q16	-1.77	.17	1.09	0.63	.46
Q17	-1.01	.12	1.08	0.88	.55
Q18	-1.02	.12	0.97	0.81	.52
Q19	1.95	.08	0.93	0.97	.73
Q20	1.60	.07	0.97	0.94	.72
Q21	1.02	.07	2.07	1.99	.56
Q22	0.58	.08	1.05	0.97	.71
Q23	0.35	.08	1.09	0.96	.69

3.2.5. Person Measure Quality

Person measure quality was investigated by examining participant measure range, size of standard errors, and fit statistics. Person measures exhibited considerable variation (Range = -.75 to 4.37 logits) and the standard error for each person measure was quite small (M = .24, SD = .07). Of the 140 participants, 6 (about 4% of the overall data set) severely overfit the RRSM and were removed from the analysis.

3.2.6. Construct Hierarchy

In psychometrics, a construct is thought of as a hierarchy in which the items rank-order themselves onto a ruler. One of the most powerful features of Rasch output is the construct map (see Figure 2). For a detailed overview of the construct map and how to interpret readers are encouraged to see Royal [11]. In brief, the map contains two sides with persons on the left and items on the right. For this particular map, items appearing at the top are considered the most severe forms of academic misconduct, whereas those at the bottom are considered the least severe forms of academic misconduct. Respondents rated item #10, *Taking a graded quiz or examination for another student*, as the most severe form of misconduct and item #19, *Failing to prepare adequately for a group assignment or laboratory*, as the least severe form of misconduct.

In an effort to investigate invariance, a Differential Item Functioning (DIF) procedure was performed to determine if the construct was stable across student subpopulations, particularly gender and class year. More specifically, measures for various student subgroups on each item were calculated and the differences were contrasted using the iterative-logit

(Rasch-Welch) method described in Linacre [17]. Because multiple comparisons were made among the 23 items, a Bonferroni adjustment was made to combat family-wise. As such, the *p*-value originally set at 0.05 was reduced to 0.002. No evidence of DIF was discernible across gender or class year.

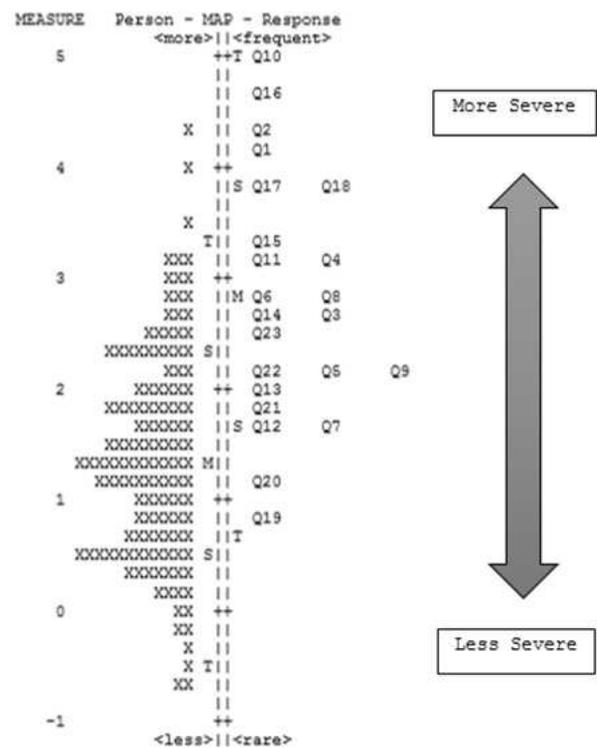


Figure 2. Construct Hierarchy.

## 4. Discussion

### 4.1. Psychometric Properties of the EAS

According to Messick [18], construct validity is the integration of any evidence that impacts the interpretation or meaning of a score. Messick described six aspects of validity: content, substantive, structural, generalizability, external and consequential. Because most measurement scholars have adopted Messick's framework for evaluating construct validity evidence, we will use his framework to evaluate the construct validity of the EAS.

A Rasch-based PCA of standardized residual correlations provided evidence of unidimensionality. This evidence supports the substantive aspect of validity. Reliability estimates between .92 and .93 indicate measures are highly reproducible, thus speaking to the generalizability aspect of validity. Evidence suggests the rating scale was appropriate and functioned appropriately given the sample of respondents. This evidence supports the structural aspect of validity and what Lopez [19] calls communicative validity. Items were evidenced to be psychometrically sound, with only one item potentially over-fitting the Rasch model's expectations. This evidence speaks to the content aspect of validity. When investigating differential item functioning (DIF), no items exhibited differential functioning across gender and class year subpopulations. This speaks to the systematic validity of the measures, and provides additional support for the generalizability aspect of validity. Because there is currently no extant literature on the perceived severity of cheating behaviors in veterinary medicine we have no basis to compare findings across studies. Therefore, we present no evidence to support the external aspect of validity. Finally, because the measures produced from the EAS have no potential repercussions for students we present no evidence of the consequential aspect of validity [20]. Collectively, validity evidence suggests there is an abundance of evidence to support the construct validity of the EAS.

### 4.2. Discussion of Substantive Findings

Generally speaking, students responded to the EAS in a manner mostly consistent with theoretical expectations. Items that were more overt and egregious, such as items about copying from other students, using unauthorized materials, changing an answer and declaring a misgrade after a score result had been rendered, were all rated as quite severe forms of misconduct. Not surprisingly, items pertaining to effort and preparation for group projects and labs were rated among the least severe forms of misconduct.

A number of useful psychological frameworks are available to help synthesize the findings. Bunn, Caudill and Gropper [21] suggest student cheating can be broken down into two distinct categories: planned cheating and panic cheating. Planned cheating involves deliberate actions and preparation, whereas panic cheating is unplanned and spontaneous. Bunn et al.'s research revealed that panic cheating is far more prevalent than planned cheating [22].

A number of researchers have discussed the notion of individuals using neutralization techniques to justify their actions and behaviors [23-24]. Sykes and Matza proposed five neutralization techniques, which include: Denial of responsibility (e.g., acts due to outside forces beyond one's control), denial of injury (e.g., actions do not hurt anyone), denial of the victim (e.g., the victim was deserving of the consequences), condemnation of the condemners (e.g., lack of respect for those who might condemn their actions) and appeal to higher loyalties (e.g., maintaining loyalty to a group as opposed to larger society). It seems certain students who appeal to higher loyalties may be likely to commit the most severe forms of misconduct, as making a sacrifice for a friend, peer, etc. may be perceived as more important than a demand from the institution that integrity be upheld. Some students are also likely to justify their actions and behaviors by convincing themselves that misconduct is a victimless crime, while others may perceive misconduct as a retaliatory measure against faculty whom they deem to be too difficult, or not particularly likeable.

Social norms theory suggests people behave in a manner consistent of their peers, but when the frequency of an action or behavior is overestimated it can result in an increase in students' behavior [25]. If students perceive academic misconduct to be an everyday part of the student culture it is likely instances of misconduct will increase simply due to these perceptions [26].

In developing strong institutional academic conduct policies, it is critical to consider students' actions and motives from some of the aforementioned frameworks. For example, how can one better regulate misconduct that occurs as a result of careful planning versus misconduct that occurs spontaneously? Knowing that some students may apply situational ethics or some variety of neutralization techniques to justify their actions or behaviors, how can we create policies that safeguard against these actions? Knowing that colleges and universities consist of many groups, each with their own unique social norms, how can we create effective one-size-fits-all academic conduct policies that cover a variety of contingencies? Some of the questions may be answered by investigating the literature on cheating, detection and prevention. However, questions pertaining to policy development and enforcement may be more difficult to answer as research often reveals inconsistent findings about the effectiveness of various policies [27].

### 4.3. Implications

We believe this research has both methodological and practical implications for faculty and researchers across an array of academic disciplines. With regard to methodological implications, this study provides a thorough psychometric validation using a state-of-the-art measurement model that many measurement experts consider to be the "gold standard" approach to such studies. The use of Messick's framework to evaluate validity evidence and make a collective judgment about construct validity is something that other researchers may find useful when attempting to discern

validity.

Practically speaking, this research provides a useful tool for others to measure the extent to which students perceive a variety of potential academic offenses as misconduct. Having insights about what exactly students perceive to be misconduct is critical, as conflicting views with faculty and college/university administration may potentially result in an unhealthy academic environment that is frustrating for all parties. Further, understanding students' perceptions of academic misconduct offenses is critical for informing and developing equitable classroom and institutional policies. More specifically, knowing what specific actions and behaviors students consider to be severe (or not) forms of academic misconduct can help institutions develop sound policies that are used to preserve academic integrity, enforce compliance, and do so in a defensible manner [27].

#### 4.4. Limitations and Future Research

As with most surveys, one potential limitation of this research was the self-reported nature of data collection. Another potential limitation was that approximately 10% fewer students from the 1<sup>st</sup> year and 4<sup>th</sup> year classes completed the survey than 2<sup>nd</sup> and 3<sup>rd</sup> year students. First and fourth-year students encounter very different experiences in veterinary education, as 4<sup>th</sup> year students are typically gaining hands-on experience on clinical rotations whereas 1<sup>st</sup> year students are still adjusting to a professional program. Future research might investigate the extent to which students nearing completion of their studies share or differ in their perspectives about misconduct from other students at different stages in the pipeline. As students tend to become more immersed in professional norms do they tend to exhibit more harsh or more lenient perspectives on academic misconduct? Future research might also compare students' perceptions across an array of related disciplines (e.g., professional health programs, arts & humanities, social sciences, etc.).

#### 4.5. Conclusions

Obtaining clear measures of what exactly constitutes academic misconduct is critically important for college and university faculty and administrators. Given the diverse generational, cultural and societal differences at play in any institution it is important that faculty routinely monitor perceptions of misconduct and continually update academic conduct policies accordingly to ensure fairness and equity for all. In order to guide the development of such policies, it is necessary to have empirical data so as to make the policy more defensible. In the present study, we evaluated the psychometric properties of the EAS and judged it to be a psychometrically-sound instrument capable of helping inform such policy decisions as there was an abundance of evidence to support construct validity. We encourage others to adopt the EAS and conduct academic misconduct surveys at their respective institutions as well.

## References

- [1] R. T. Burrus, K. M. McGoldrick, and P. W. Schuhmann. "Self-reports of student cheating: Does a definition of cheating matter?," *Journal of Economic Education*, 38(1), 3-16, 2007.
- [2] G. J. Cizek, *Detecting and preventing classroom cheating: Promoting integrity in assessment*. Thousand Oaks, CA, Sage, 2003
- [3] E. B. Stern, and L. Havlicek. "Academic misconduct: Results of faculty and undergraduate student surveys," *Journal of Allied Health*, 15(2), 129-143, 1986.
- [4] T. O. Bisping, H. Patron, and K. Roskelley. "Modeling academic dishonesty: The role of student perceptions and misconduct type," *Journal of Economic Education*, 39(1), 4-21, 2008.
- [5] K. D. Royal, J. V. Parrent, and R. P. Clark. "Measuring education majors' perceptions of academic misconduct: An item response theory perspective," *International Journal for Educational Integrity*, 7(1), 18-29, 2011.
- [6] T. G. Bond and C. M. Fox. *Applying the Rasch Model. Fundamental measurement in the human sciences*, 2<sup>nd</sup> edition. Lawrence Erlbaum Associate, 2007.
- [7] T. Salzberger. "The illusion of measurement: Rasch versus 2-PL," *Rasch Measurement Transactions*, 16(2), p. 882, 2002.
- [8] B. D. Wright. "Fundamental measurement," *Rasch Measurement Transactions*, 11(2), p. 558, 1997.
- [9] B. D. Wright. *Measurement for Social Science and Education: History of Social Science Measurement*. MESA Memo #62, Available at: <http://www.rasch.org/memo62.htm>, 2007.
- [10] G. Engelhard, Jr. *Invariant measurement: Rasch measurement in the social, behavioral and health sciences*. Routledge, 2013.
- [11] K. D. Royal. "Making meaningful measurement in survey research: A demonstration of the utility of the Rasch model," *IR Applications*, 28, 1-16, 2010.
- [12] D. Andrich. "A rating formulation for ordered response categories," *Psychometrika*, 43, 561-573, 1978.
- [13] L. M. Linacre. WINSTEPS® (Version 3.90.0). Computer Software. Beaverton, OR: Winsteps.com, 2015.
- [14] B. D. Wright and G. N. Master. *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press, 1982.
- [15] J. M. Linacre. "Optimizing rating scale category effectiveness," *Journal of Applied Measurement*, 3(1), 85-106, 2002.
- [16] B. D. Wright and J. M. Linacre. "Reasonable mean-square fit values," *Rasch Measurement Transactions*, 8, 370, 1994.
- [17] J. M. Linacre. *Differential item functioning DIF pairwise*. Available at: [http://www.winsteps.com/winman/table30\\_1.htm](http://www.winsteps.com/winman/table30_1.htm), 2015.
- [18] S. Messick. "Validity," In R. L. Linn (Ed.) *Educational Measurement* (3<sup>rd</sup> ed., pp. 13-103). New York: Macmillan, 1989.

- [19] W. Lopez. "Communication validity and rating scales," *Rasch Measurement Transactions*, 10(1), 482-483, 1996.
- [20] K. D. Royal and J. C. Puffer. "The consequential validity of ABFM examinations," *Journal of the American Board of Family Medicine*, 27(3), 430-431, 2014.
- [21] D. N. Bunn, S. B. Caudill, and D. M. Gropper. "Crime in the classroom: An economic analysis of undergraduate student cheating behavior," *The Journal of Economic Education*, 23(3), 197-207, 1992.
- [22] T. C. Grijalva, C. Nowell, and J. Kerkvliet, J. "Academic honesty and online courses," *College Student Journal*, 40(1), 180-185, 2006.
- [23] E. E. LaBeff, R. E. Clark, V. J. Haines, and G. M. Dickhoff. "Situational ethics and college student cheating," *Sociological Inquiry*, 60(2), 190-198, 1990.
- [24] G. M. Sykes and D. Matza, D. "Techniques of neutralization: A theory of delinquency," *American Sociological Review*, 22(6), 664-670, 1957.
- [25] S. F. Hard, J. M. Conway, and A. C. Moran. "Faculty and college student beliefs about the frequency of student academic misconduct," *The Journal of Higher Education*, 77(6), 1058-1080, 2006.
- [26] D. L. McCabe, L. K. Trevino and K. D. Butterfield. "Academic integrity in honor code and non-honor code environments: A qualitative investigation," *The Journal of Higher Education*, 70(2), 211-234, 1999.
- [27] G. J. Cizek. *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum, 1999.