Measuring education majors' perceptions of academic misconduct: An item response theory perspective

Kenneth D. Royal
College of Education, University of Kentucky
kdroya2@uky.edu

Jonathon V. Parrent
Department of Educational Leadership Studies
University of Kentucky
jay.parrent@kctcs.edu

Ryan P. Clark
Department of Educational Leadership Studies
University of Kentucky
ryan.clark@nelson.kyschools.us

Abstract

The purpose of this study was to construct a psychometric ruler that illustrates university undergraduate education majors' perceptions of academic misconduct. A survey consisting of 38 items that pertain to issues of academic misconduct were administered to an undergraduate sample at a large state university. Utilising Rasch measurement analyses to construct objective measures, students' responses were modeled along a truly linear and equal interval continuum to produce a hierarchy of perceived academic offenses. Results and policy implications are discussed.

Introduction

Maintaining academic integrity is critical for all stakeholders in higher education. For students, this is particularly important because the eventual degree obtained should carry some measure of value and respect. When students engage in academic misconduct they compromise not only their personal integrity, but also that of the courses for which they are enrolled, the degree programs for which they belong, and the institution for which they inhabit. Of course, not all forms of academic misconduct are created equal. It is easy to understand that some offenses may be considered more severe than others. Unfortunately, these lines of academic misconduct are often grey and it is difficult to provide objective measures that truly delineate these degrees of severity.

The purpose of this study was to construct a psychometric ruler that illustrates undergraduate education majors' perceptions of academic misconduct. Although the literature on academic misconduct is vast and rather comprehensive, no study to date has utilised a powerful state-of-the-art Item Response Theory (IRT) technique,

namely Rasch measurement, to analyse data. Rasch measurement models are the only psychometric methods available that possess the properties necessary for objective measurement comparable to that of the physical sciences (Rasch, 1960; Wright, 1967; Wright, 1999; Wright & Douglas, 1986). Most traditional statistical techniques erroneously treat raw scores as measures, ordinal scales as interval, and assume that all items are of equal importance. These shortcomings can have significant consequences with regard to constructing valid measures and the inferences consumers make based on these results. This research overcomes the aforementioned weaknesses of traditional statistical methods and provides measures of students' perceptions of academic misconduct that are both more informative and meaningful, and possibly more valid.

The present study will begin by providing an overview of selected literature on the topic of academic misconduct. This will be followed by a brief introduction to Rasch modeling. The research design and methodology will then be presented, and results of the data analysis will be provided. A discussion of the results and implications will conclude the study.

Review of selected literature

College and university faculty and administrators in the United States (US) face a variety of student behaviours that can be deemed 'academic misconduct'. We define academic misconduct as any behaviour that is unethical or lacks integrity. Academic misconduct is a growing concern on most college and university campuses and technological improvements continue to pose additional threats. Students now have a number of electronic resources that make committing acts of academic dishonesty easier, faster, and more difficult for faculty to uncover (Szabo & Underwood, 2004). Term papers are available to purchase online on any conceivable topic and cell phone technology literally places the answers at students' fingertips. But the problem of academic misconduct is not a new one. In the last three decades, it has been reported that 13% to 95% of students engage in some form of academic misconduct in higher education (McCabe & Trevino, 1997). This is particularly disturbing as academic dishonesty has the potential to disrupt and threaten the academic process (Simon et al., 2004).

Students and university faculty perceptions and experiences support the notion that misconduct on campus is a problem and students engage in these behaviours frequently (Schmelkin, Kaufman, & Liebling, 2001; Engler, Landau, & Epstain, 2008). Besvinick (1983) stated that self-interested and amoral students that are enrolled in higher education may threaten the integrity of the institution. Therefore, understanding and combating the problem of academic misconduct is essential to maintaining academic integrity at colleges and universities.

Understanding the magnitude of the problem and all the factors that may influence and/ or better define the problem is the first step needed to combat this issue. A study conducted by McCabe and Trevino (1997) provided valuable insights about individuals and contextual issues regarding whom and in what context academic misconduct might be more prevalent. Results found that individuals from certain demographics, particularly males and younger students, were more likely to engage in academic misconduct. Also, students who were involved in extracurricular activities were also more inclined to engage in academically dishonest behaviours. Conversely, students who maintained a higher grade point average (GPA) and were considered academically successful were less likely to engage in academically dishonest acts.

Contextual issues appeared to have an impact on whether or not a student might take part in an academically dishonest act as well. Peer influence in the form of both actions and opinions may contribute to academic misconduct. Particularly, if one's peers are

more approving of dishonest and unethical academic behaviours, a student is potentially at a higher risk for committing academically dishonest acts as well. Consequences/penalties were also significant factors. Students who believed faculty took academic misconduct seriously and enforced policies that carried severe penalties were not as likely to risk taking part in academically dishonest acts.

Although we may know who is likely to engage in academic misconduct, the lack of a clear boundary establishing what truly constitutes academic misconduct remains a concern. No established or generally accepted definition of academic misconduct exists in the literature (Schmelkin, Gilbert, Spencer, Pincus, & Silva, 2008), but an acceptable definition must be created (Maramark & Maline, 1993). Multiple studies have been conducted to establish how students and faculty perceive behaviours that could be classified as academic misconduct (see Bisping, Patron, & Roskelley, 2008; Levy & Rakovski, 2006; Schmelkin et al., 2008; and Schmelkin et al., 2001). From these studies, it appears these perceptions on the severity of behaviours constituting academic misconduct rarely align (Schmelkin et al., 2008; Schmelkin, et al., 2001). This difference is significant because, "if a student believes that the behavior constitutes misconduct, he or she is less prone to undertake it" (Bisping et al., 2008, p. 11).

With the aforementioned studies in mind, our paper is rather unique in that it places a premium on psychometric methods. Often times, IRT techniques provide results that are different than those derived from traditional statistical analyses, despite analysing the exact same dataset. The following section provides a brief introduction to the methodology employed in this study and makes an argument for a greater usage of these powerful techniques in the academic misconduct literature.

Rasch measurement
Psychometrics is the field of study that attempts to measure psychological constructs, particularly latent traits such as knowledge, abilities, attitudes, personality traits, and so on. In the social and behavioural sciences, a construct is thought of as a hierarchy. For example, consider a standardised test. A good test consists of items of varying degrees of difficulty. Likewise, those who take the test possess varying degrees of ability. In order to obtain any meaningful information about what a test-taker can do (or knows), we must understand exactly how difficult each item is and how able (or knowledgeable) each test-taker is. A more able person should always have a higher probability of getting an item correct than someone who is less able. Likewise, a more difficult item should always have a greater probability of being answered incorrectly than an easier item.

The problem with traditional statistical methods is they do not account for such differences. In fact, they are 'circular dependent', meaning the sample of participants and the items are inherently dependent upon one another. In the 1950s, psychometricians sought to overcome the deficiencies of traditional statistical methods (also known as 'Classical Test Theory'), and IRT methods were developed. These methods are regularly employed in high-stakes testing today because of the strong theoretical underpinnings and their powerful capabilities. However, these powerful models are not restricted only to tests; these techniques can be extended to surveys and other scenarios as well (see Bond & Fox, 2007; Mueller & Bradley, 2009; Royal, 2010; Smith & Smith, 2004; Wolfe, Ray, & Harris, 2004). As mentioned previously, the latent trait measured in a test is typically one's ability. In a survey scenario, the latent trait measured is typically one's propensity to endorse a particular item, otherwise known in the measurement literature as 'endorsability'.

Rasch measurement models have routinely been used since the 1960s to produce objective, linear measures of various phenomena. Rasch models are logistic, latent trait models of probability for monotonically increasing functions. Unlike statistical models that are developed to fit data, Rasch measurement models are static wherein data are

fit to the models. Rasch models assume the probability of a respondent endorsing a particular item is a logistic function of the relative distance between the person and item location on a linear continuum. The present study utilised the Rating Scale Model (Andrich, 1978). The formulae for the Rating Scale Model are presented below:

$$\ln (P_{nij}/P_{ni\ (j-1)}) = B_n - D_i - F_j$$

 where,

P_{nij} is the probability that person n encountering item $_i$ is observed in category $_j$, B_n is the 'ability' measure of person $_n$, D_i is the 'difficulty' measure of item $_i$, the point where the highest and lowest categories of the item are equally probable. F_j is the 'calibration' measure of category $_j$ relative to category $_{j-1}$, the point where categories $_{j-1}$ and $_j$ are equally probable relative to the measure of the item. No constraints are placed on the possible values of F_j.

Although the process of Rasch analysis is well documented in the literature (see Wright & Stone, 1979; Wright & Stone, 1999; Smith, Jr. & Smith, 2004; Bond & Fox, 2007), it would suffice to say that the analysis is largely concerned with the extent to which observed data match what is expected by the model. Additionally, any time one performs a Rasch analysis numerous quality control checks must be performed. These quality control checks essentially provide a validation that the measures are of sufficient quality. Such checks consist of evaluating overall data to model fit, person and item level fit indices, rating scale functioning, reliability and separation indicators, measure stability based on spread of item variances and size of standard errors, and so on. If sufficient evidence exists that quality measurement is taking place one can have a great deal of confidence in the truthfulness of the results. Although the quality control (or validation) process is rather lengthy, it is typically necessary to report. Provided the purpose of this study is largely to illustrate an application of Rasch measurement, the authors opted to report much of the technical specifics so that other researchers interested in this technique may better understand their output.

Methodology

Instrumentation and data

In 2008, Bisping et al. published a study that modelled undergraduate students' perceptions of academic misconduct. Bisping et al.'s instrument contained three parts: 1) a section containing demographic items; 2) a section which contained 38 items asking whether various situations constituted academic misconduct, and whether students had engaged in each of these actions while in college; and 3) a section with 20 items that asked students about the extent to which they believe a series of factors would affect the frequency of academic misconduct. The instrument used in Bisping et al.'s study was a modified version of that used by Stern and Havlicek (1986). For the present study, only the second section of Bisping et al.'s instrument was used to measure students' perceptions of academic misconduct. A modification was made to the rating scale as well. Bisping et al.'s rating scale consisted of a range from 1 to 4. For the current study, a seven-point semantic differential scale was provided with a rating of 1 indicating 'Not academic misconduct', and a 7 indicating 'Severe academic misconduct'. This modification was made to improve measurement precision.

A survey was administered to six sections of an undergraduate education course at a large southeastern university in the US. The education course was selected because it was a core requirement for all education majors and was the most heavily enrolled course within the college. The census sampling approach yielded a total of 114 responses, with a response rate of nearly 100%.

Results

As mentioned previously, whenever Rasch analyses are employed, a series of quality control checks are conducted. This is essentially a validation process that confirms that all the elements are functioning properly. Psychometric indicators identify where problems might exist and allow researchers the opportunity to revisit potential issues or otherwise press forward. We begin by investigating overall summary statistics.

Summary statistics illustrate important information about the quality of measurement. Particularly, summary statistics provide information about average person and item measures (accompanied with standard deviations), overall model error, information about the reliability of the measures produced, and information about how well the data fit the Rasch model. Table 1 provides a summary of these data, while Table 2 provides a more thorough breakdown of reliability estimation.

Table 1:
Summary statistics

	Measure	Model Error	Infit Mean Square	Outfit Mean Square
Students				
Mean	0.06	0.14	1.04	1.02
S.D.	0.58	0.01	0.44	0.43
Items				
Mean	0.00	0.09	1.04	1.02
S.D.	1.02	0.03	0.27	0.26

Table 2.
Reliability and Separation

	Reliability (Real)	Reliability (Model)	Separation (Real)	Separation (Model)
Students	0.93	0.95	3.79	4.17
Items	0.99	0.99	10.54	11.11

With Rasch measurement, it is essential that the data fit the model well, as this is paramount for establishing internal construct validity. Fit statistics are key indicators, with values of 1.00 indicating perfect fit. With Rasch measurement, it is common practice to evaluate misfitting persons and items to determine if any of the measures are so grossly misfitting that they negatively affect the precision of measurement. Criteria for removal in survey research typically involve persons and items misfitting at levels below 0.4 or above 1.6 (Wright & Linacre, 1994) for either infit or outfit mean square values. Once misfitting persons and/or items are identified, the researcher must make a decision to keep or disregard these data. In this study, 11 persons grossly misfit the Rating Scale Model's expectations and degraded measurement precision. Therefore, these individuals were removed from the analysis. Upon removing these misfitting persons, data-to-model fit was quite good.

Separation is the ratio of sample deviation, corrected for error, to the average estimation error (Linacre, 2010). Another way to think of it is the spread of the sample into various statistically distinguishable levels. In Rasch measurement, reliability is reported as both Real and Model. Real reliability refers to lower bound of reliability, while model reliability refers to the upper bound. That is, real reliability is the 'worst case scenario' estimate and model reliability is the 'best case scenario' estimate. Here, person estimates of endorsability fall between 0.93 and 0.95, and item estimates are stable at 0.99. These estimates indicate there was sufficient variance in both the sample of persons and the pool of items and enough items with sufficient discrimination to spread out person ratings. Further, rating scale categories were

sufficient for each item and were fully utilised by respondents. Additionally, items were reasonably well targeted to the audience of undergraduate students based on the comparison of person and item average measures. High reliability estimates also provide some support for the generalisability aspect of validity.

Item statistics
Item statistics provide estimates for both the difficulty of each item, as well as its associated standard error. Table 3 identifies the difficulty measure and standard error for each of the 38 items. Infit statistics focus on centralised performance and outfit statistics are sensitive to outliers and unexpected responses at the extreme ends of the scale. Item infit and outfit mean square statistics ranged from 0.60 to 1.40, all within the acceptable range of quality measures as suggested by Wright and Linacre (1994). The confirmation of item fit provides evidence of item quality and content validity.

Table 3:
Item measures and standard error

Item		Measure	Error
Q1	Copying from others during tests	-1.31	0.10
Q2	Preparing cheat sheets but not using them	0.37	0.07
Q3	Using cheat sheets during tests	-1.56	0.12
Q4	Having someone else do your homework	-1.42	0.11
Q5	Copying or buying a paper but presenting it as your own	-2.04	0.15
Q6	Copying from a source without giving proper reference	-0.46	0.08
Q7	Turning in the same paper in more than one class	0.53	0.07
Q8	Asking other students about the content of a test you are about to take	0.29	0.07
Q9	Using the work of other students in your paper without giving them credit	-0.26	0.07
Q10	Reading Cliffs Notes or condensed versions instead of full-length assignments	1.26	0.09
Q11	Listing unread material in the references of a paper	0.28	0.07
Q12	"Making up" references in a paper	-0.37	0.07
Q13	Listing non-pertinent material in the references of a paper	0.59	0.07
Q14	Working in groups when assignments are meant to be completed individually	0.89	0.08
Q15	"Sitting for" a student during a test	-2.59	0.20
Q16	Memorizing test questions to review them later	1.15	0.09
Q17	Having a paper edited, when style, etc. are not graded	1.80	0.12
Q18	Having a paper edited, when style, etc. are being graded	1.19	0.09
Q19	Giving false excuse to miss class or postpone assignments	0.09	0.07
Q20	Keeping a page or multiple pages of a test when the teacher does not allow it	-0.96	0.09
Q21	Getting a copy of a test by having a student "sit in" on the test and not turn it in	-1.25	0.10
Q22	Removing reserved material from a file to prevent other students from viewing it	-0.54	0.08
Q23	Looking at other's test and keeping your answer when both responses are the same	-0.22	0.07
Q24	Changing your answer on a test after it was graded and reporting a false misgrade	-1.40	0.11
Q25	Purchasing or being given notes from a fellow student	1.16	0.09
Q26	Marking two answers on a computer answer sheet when only one is allowed	0.42	0.07
Q27	Marking two answers on a hand scored test, making your choice unclear	0.13	0.07
Q28	Permitting others to look at your test	-0.71	0.08
Q29	Falsely claiming to have handed in a test or a paper	-1.02	0.09
Q30	Doing less than your fair share in a group project	-0.11	0.07
Q31	Trying to bias a professor	0.51	0.07
Q32	Changing margins or format of a paper to make it appear longer or shorter	0.26	0.07
Q33	Talking with classmates during lectures	0.92	0.08
Q34	Watching movies or playing games on a laptop during class	0.75	0.08
Q35	Reading or studying non-related class material during lectures	1.04	0.08
Q36	Arriving late to class	0.93	0.08
Q37	Leaving class early	0.83	0.08
Q38	Leaving your cell phone on during lectures or examinations	0.86	0.08

Rating scale quality

To investigate the quality and structural validity of the seven-point rating scale used in this survey, rating scale diagnostics were evaluated. Rating scale diagnostics provide counts and percents for each response option, indicating the extent to which each option was utilised by respondents. Infit and outfit mean square statistics indicate the fit of each response to the structure of the rating scale. Structure calibration refers to the calibrated measure of transition between categories. Also called 'step calibration', this measure indicates how difficult it is to observe each category. Table 4 indicates survey respondents utilised the full range of the scale, with the least utilised rating category utilised 12% of the time, and the most utilised rating category utilised 19% of the time. Fit statistics reveal all the response options fit the scale and none appear problematic.

Table 4:
Rating scale effectiveness

Category	Count	Percent	INFIT MNSQ	OUTFIT MNSQ	Structure Calibration	Category Measure
1	733	19%	1.05	1.07	NONE	2.13
2	462	12%	0.96	0.92	-0.52	1.05
3	460	12%	1.02	0.97	-0.62	-0.48
4	513	13%	0.96	0.95	-0.38	-0.04
5	478	12%	1.00	0.99	0.19	0.42
6	498	13%	1.02	0.98	0.56	1.06
7	751	19%	1.16	1.18	0.77	2.28

Demographics

Demographic information is provided for the 114 persons who responded to the survey. The majority of the sample were female (76%), with most students ranging in age from 20 to 24. There was not a great deal of racial diversity in the sample, as approximately 90% reported being white. The courses were comprised of predominantly junior (third-year) classmen, although there were also a number of sophomores (second-year) and seniors (fourth-year) as well. Most students possessed a GPA greater than 3.0 on the 4.0 scale.

Table 5:
Personal demographic characteristics of student respondents (n = 114)
Psychometric ruler

Characteristic	n	%
Sex		
Male	26	22.8
Female	87	76.3
Age		
18-19	8	7.1%
20-21	75	66.4%
22-24	24	21.2%
25-29	4	3.5%
30-34	0	0.0%
35-39	1	0.9%
Over 40	1	0.9%
Race		
American Indian/Alaskan Native	1	0.9%
Black or African American	6	5.3%
Two or More Races	2	1.8%
White	102	89.5%
Year		
Sophomore	19	16.8%
Junior	60	53.1%
Senior	33	29.2%
Graduate Student	1	0.9%
Cumulative GPA		
Between 2.5 and 2.99	18	16.1%
Between 3.0 and 3.49	48	42.9%
Between 3.5 and 3.99	40	35.7%
4.0	6	5.4%

The psychometric ruler, also known as an 'item map' is presented in Figure 1. This ruler illustrates the construct hierarchy of items with regard to how easy/difficult each are to endorse. To interpret the map, persons appear on the left side of the map, and items appear on the right. Indicators of 'M', 'S', and 'T' along the vertical axis indicate the mean, standard deviation and two standard deviations for both people and items, respectively. The distribution of students appears normally distributed, and the items spread the length of the continuum indicating they are well-targeted to this audience. Items appearing at the top of the map are the most difficult items to endorse (or agree with). Conversely, items appearing at the bottom of the map are the easiest items to endorse. Based on the item hierarchy, it appears item *Q17* is the most difficult item and *Q15* is the easiest item for survey respondents to endorse.

```
                STUDENT - MAP - RESPONSE
                    <more>|<rare>
       2                  +T          Most difficult to endorse
                           |  Q17
                       .   |
                           |
                           |  Q10
                   .#   T|  Q16    Q18    Q25
       1           .#   +S  Q33    Q35    Q36
                   ##    |  Q14    Q37    Q38
              .####  S|  Q13    Q34
               .###    |  Q26    Q31    Q7
               .####   |  Q11    Q2     Q32    Q8
             .#######  |  Q19    Q27
       0    ######### M+M
               .###    |  Q23    Q30
              .#####   |  Q12    Q9
                ## S|  Q22    Q6
               ###    |  Q28
                ##    |
      -1           .  +S  Q20    Q29
                   .  T|
                   .   |  Q1     Q21    Q24    Q4
                       |  Q3
                       |
                       |
      -2              +T  Q5
                   .   |
                       |
                       |
                       |  Q15
                       |
      -3              +            Least difficult to endorse
                    <less>|<frequently>
       EACH "#" IS 2. EACH "." IS 1.
```
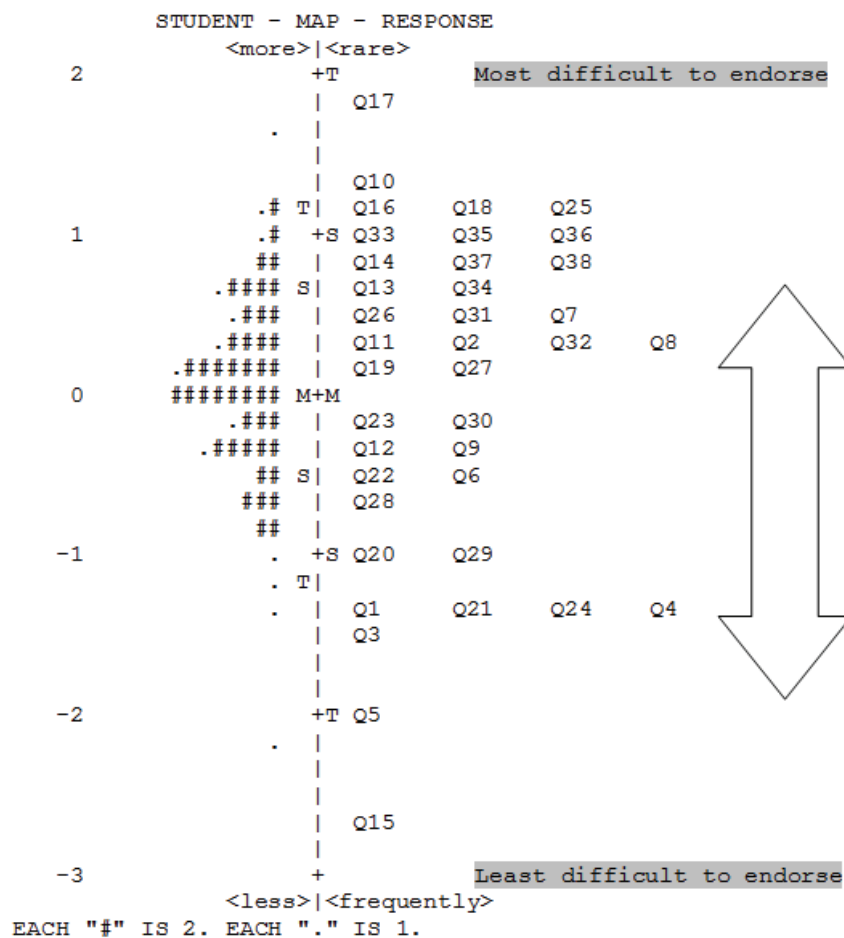
Figure 1. Item map

Discussion

When asked to consider the severity of various items as it pertains to academic misconduct, it is not surprising that students identified item *Q15 'Sitting for' a student during a test* as the easiest to endorse. Stated another way, this item was considered to be the most severe form of academic misconduct among the 38 items presented on the survey. Other items receiving strong endorsement from students' as being a severe form of misconduct include *Q5 Copying or buying a paper but presenting it as your own* and *Q3 Using cheat sheet during tests*. In fact, *Q15* was far easier for students to endorse than *Q5* and *Q3*. The psychometric distance between these items may be somewhat surprising given the explicit nature of these offenses.

Across the spectrum, behaviours traditionally viewed as misconduct or certainly ethically challenged did not register with many students. For example, students perceived *Q12 'Making up' references in a paper* as less severe than *Q22 Removing reserved material from a file to prevent other students from viewing it*. Additionally, items such as *Q2 Preparing cheat sheets but not using them*, *Q32 Changing margins or formatting to make a paper appear longer or shorter*, and *Q11 Listing unread material in the references of a paper* are generally perceived as a greater form of academic misconduct than *Q26 Marking two answers when only one is allowed*, *Q31 Trying to bias a professor* and *Q7 Turning in the same paper in more than one class*. Another interesting finding is that classroom behavioural items (*Q33*, *Q35*, *Q36*, *Q37* and *Q8*) that involve arriving late, leaving early, leaving cell phone on, talking with classmates, etc., tend to be regarded as a more severe form of academic misconduct

than *Q10 Reading Cliffs Notes or condensed versions of full-length assignments* and *Q25 Purchasing or being given notes from a fellow student*.

It is evident in many of these findings that different types of ethical issues are involved. Some types of issues are clearly easier to identify than others. The psychometric ruler is particularly helpful in allowing easy identification of themes in the interpretation of results. For instance, by examining the ruler one can see that items such as *Q38 Leaving your cell phone on during examinations* and *Q33 Talking with classmates during lectures* appear very close to one another on the continuum. Likewise, *Q36 Arriving late to class* and *Q37 Leaving class early* are also located within a very close proximity to one another. There is no doubt a conceptual relationship exists between these items, as each illustrates a facet of respectful classroom behaviour. The ability to illustrate these concepts and identify thematic items can provide a great deal of insights about the presence of constructs, thus providing an alternative (or perhaps complimentary technique) to correlational and factor analyses.

It should be noted that these results are not definitive. The item hierarchy might change depending on the demographics of the sample. In this study, the sample consisted of almost entirely juniors, with the vast majority being female. It is possible that the item hierarchy might look different if the sample contained students from other demographics, class levels and/or institutional types. Future research should attempt to replicate this study on different samples to determine whether the construct hierarchy remains intact across undergraduate student populations.

An interesting corollary study would be to administer this survey instrument to college/ university faculty to collect comparative data. A comparison of perceptions of academic misconduct could be particularly useful in determining the extent to which students and faculty views correlate or deviate from one another. This has significant implications for policy, as honour codes, disciplinary measures and other issues in the realm of academic integrity are generally created, or at the very least driven, by faculty. Instances where there is a significant disconnect between faculty and students with regard to what is considered inappropriate could serve as a useful opportunity to revise or revisit potentially dated policy. Also, identifying faculty and student disconnect in this area could also provide an opportunity for faculty and administrators to more explicitly inform students of what is considered inappropriate at one's respective campus.

Conclusions

This research produced a psychometric ruler that illustrates US university undergraduate education majors' perceptions of academic misconduct. By utilising Rasch measurement analyses to construct linear and objective measures, students' survey responses were modelled along a linear, equal interval continuum comparable to a ruler used in the physical sciences. The resulting psychometric ruler provides a unique and perhaps more meaningful way to interpret quantitative results, as the interactions between the person latent trait (endorsability) and the item's difficulty to endorse can be directly compared. Substantive findings, avenues for future research, and other implications were also provided. Policy implications that should be of significant interest to all teaching faculty and college and university administrators were also addressed.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.

Besvinick, S. (1983). Integrity and the future of the university. *Journal of Higher Education, 54*(5), 566–573.

Bisping, T. O., Patron, H., & Roskelley, K. (2008). Modeling academic dishonesty: The role of student perceptions and misconduct type. *Journal of Economic Education, 39*(1), 4–21.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model. Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associate.

Engler, J. N., Landau, J. D., & Epstein, M. (2008). Keeping up with the Joneses: Students' perceptions of academically dishonest behavior. *Teaching of Psychology, 35*(2), 99–102.

Linacre, J. M. (2010). Winsteps® (Version 3.69.1) [Computer Software]. Beaverton, Oregon: Winsteps.com.

Levy, E. S. & Rakovski, C. C. (2006). Academic dishonesty: A zero tolerance professor and student registration choices. *Research in Higher Education, 47*(6), 735–754.

Maramark, S., & Malin, M. (1993). *Academic dishonesty among college students.* (ERIC Document Reproduction Service No. ED360903).

McCabe, D., & Trevino, L. (1997). Individual and contextual influences on academic dishonesty. *Research in Higher Education*, *38*(3), 379–396.

Mueller, C. E., & Bradley, K. D. (2009). Utilizing the Rasch model to develop and evaluate items for the tacit knowledge inventory for superintendents (TKIS). *International Journal of Knowledge Management, 5*(3), 73–93*.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Royal, K. D. (2010). Evaluating faculty perceptions of student learning outcomes: A Rasch measurement analysis. *Journal of MultiDisciplinary Evaluation, 6*(14), 18–31.

Schmelkin, L., Gilbert, K., Spencer, K., Pincus, H., & Silva, R. (2008). A multidimensional scaling of college students' perceptions of academic dishonesty. *The Journal of Higher Education*, *79*(5), 587–602.

Schmelkin, L., Kaufman, A., & Liebling, D. (2001). Faculty assessments of the clarity and prevalence of academic dishonesty. Paper presented at the *American Psychological Association,*San Francisco.

Simon, C., Carr, J., McCullough, S., Morgan, S., Oleson, T., & Ressel, M. (2004). Gender, student perceptions, institutional commitments and academic dishonesty: Who reports in academic dishonesty cases? *Assessment & Evaluation in Higher Education*, *29*(1), 7590.

Smith, E. V., Jr., & Smith, R. M. (2004). *Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press.

Stern, E. B., & Havlicek, L. (1986). Academic misconduct: Results of faculty and undergraduate student surveys. *Journal of Allied Health*, *15*(2), 129–143.

Szabo, A., & Underwood, J. (2004). Cybercheats: Is information and communication technology fuelling academic dishonesty? *Active Learning in Higher Education, 5*(2), 180–199.

Wolfe, E. W., Ray, L. M., & Harris, D. C. (2004). A Rasch analysis of three measures of teacher perceptions generated from the School and Staffing Survey. *Educational and Psychological Measurement, 64*(5), 842–860.

Wright, B. D. (1967). Sample-free test calibration and person measurement. MESA Memo #1. Retrieved April 4, 2011, from www.rasch.org/memo1.htm.

Wright B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wright. B. D., & Douglas, G. A. (1986). The Rating Scale Model for objective measurement. MESA Memo #35. Retrieved April 4, 2011, from www.rasch.org/memo35.htm.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1999). *Measurement Essentials* (2nd ed.). Wilmington, DE: Wide Range, Inc.

About the authors

Kenneth D. Royal is a psychometrician at the American Board of Family Medicine and an Adjunct Professor in the College of Education at the University of Kentucky. His research interests focus on measurement, assessment, evaluation, research methods and higher education. He is currently an officer in the Rasch Measurement Special Interest Group (SIG) of the American Educational Research Association (AERA) and widely publishes research in fields such as medicine, higher education, criminal justice and psychology.

Jonathan V. Parrent is the Dean of Student Affairs at Madisonville Community College. He is currently a doctoral student at the University of Kentucky in the Department of Educational Leadership Studies and a current member of the Kentucky Board of Education. His research interests include institutional research, assessment and various issues within the higher education arena.

Ryan P. Clark is the principal at Old Kentucky Home Middle School. He is currently a doctoral student at the University of Kentucky in the Department of Educational Leadership Studies. His research interests include qualitative research, educational policy and elementary/middle school education.