

RESPONSIBLE CONDUCT OF RESEARCH MEASURE: INITIAL DEVELOPMENT AND PILOT STUDY

By: [Kelly L. Wester, PhD.](#), [John T. Willse, Psy.D.](#), and Mark S. Davis, PhD.

Wester, K. L., Willse, J., & Davis, M. S. (2008). Responsible Conduct of Research Measure: The initial development and pilot study. *Accountability in Research*, 15, 87-104.

Made available courtesy of Taylor and Francis: <http://www.tandf.co.uk/>

*****Reprinted with permission. No further reproduction is authorized without written permission from Taylor and Francis. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.*****

Abstract:

Although much discussion has been focused on research misconduct (RM) and questionable research practices, to date no self-report measures exist to examine this phenomenon. To help fill this void, the authors developed the Responsible Conduct of Research Measure (RCRM) through multiple pilot study waves involving researchers in the social and behavioral sciences. Preliminary results reveal adequate validity and reliability. The authors discuss limitations of the study as well as some possible directions for future research on this topic.

Keywords: ethics, integrity, research misconduct, responsible conduct of research

Article:

For academic scientists, research comprises a large portion of their workload. A general expectation is that research has been conducted competently and ethically. Although safeguards (e.g., institutional review boards) have been put into place to ensure that at least some aspects of research are conducted with integrity, the primary responsibility remains with individual researchers to follow established ethical guidelines for conducting and reporting research. However, very little is known regarding individuals' research behaviors outside of those who have been reported to their host institutions or federal agencies responsible for investigating research misconduct (RM). One of the reasons for this void is the lack of an appropriate instrument to measure RM and other questionable research practices (QRPs), the latter defined as departing from the acceptable practice of the relevant research community (Steneck, 2003) or from the responsible conduct of research areas provided by the Office of Research Integrity (ORI, 2000). It is this void in knowledge that led the authors to undertake the current study. Specifically, the purpose of the current study was to create an instrument that measures QRPs and RM among researchers.

Impact of RM and QRPs

RM has been estimated to occur in one out of 100 researchers (Steneck, 2006), with up to 44% of students and 50% of faculty reporting being exposed to at least one type of RM (Shafir and Kennedy, 1998). Each year, the ORI makes findings of scientific misconduct against a number of researchers, some of whom are affiliated with the most prestigious research institutions in the U.S. Recently, a record number of institutions reported a record number of RM cases (ORI, 2006).

The prevalence of RM does not speak to the impact that these behaviors have. According to Gardner (2006), known cases of RM are responsible for approximately \$427,000 of lost money per year (based on the average NIH grant); however, he proposed an estimate of the cost of RM in the billions, not including possible negative health effects or even deaths of human research participants or patients visiting physicians' offices.

Serious RM has a profoundly negative impact on the research community, professionals, and the general public; this also may be true of QRPs. Little is known about QRPs, including prevalence and impact. Steneck (2003) suggested that the occurrence of QRPs in the broader research community ranges from 10% to 40%, and that each QRP runs counter to well-established rules designated by each professional community, ultimately

compromising research integrity. Davis et al. (In press) found that between 2% and 23% of counseling professionals self-reported that they would be likely to engage in behaviors that exemplify QRPs. Based on this and other studies, it is likely that QRPs are more prevalent than RM. Consequently, it has been suggested that, due to greater volume, QRPs actually may be more damaging to the research community than outright RM (Alberts and Shine, 1994; de Vries et al., 2006).

RM and QRP Measures

It is important that we begin to identify factors related to departures from the responsible conduct of research (RCR; defined as adherence to rules, regulations or commonly accepted professional codes or standards and high moral principles; ORI, 2003; Steneck, 2006). The difficulty in identifying these factors lies in the lack of appropriate measures of RM and QRPs. As noted by the Institute of Medicine (2002) “No established measure for assessing integrity in the research community exists.” Some studies of RM have been conducted since this statement was made, but limits to the measures exist. In studies of research integrity, the primary approaches have been to examine respondent case files at ORI (e.g., Davis et al., 2001), ask researchers to select causes of misconduct from a checklist or report the number of observed incidents (e.g., Anderson et al., 1994; Anderson, 1996), or use qualitative methods to determine why participants believe RM occurs (Wiles et al., 2006).

Although these studies provide essential and important information in the area of RM, they have certain limitations. For example, qualitative research results (e.g., Davis et al., 2001; Wiles et al., 2006) do not permit inferences about the larger population of researchers. Most quantitative studies have asked individuals about their exposure to RM, as well as reasons why they believe it might occur (e.g., Anderson et al., 1994; Anderson, 1996; Broome et al., 2006; Rankin and Esteves, 1997), offering only indirect evidence of causation. Opinions of survey respondents about the behavior of others, even if reasonably informed, may not coincide with the actual causes of RM or may not yield accurate numbers or prevalence. In fact, individuals’ opinions about causation have been shown to have weak validity (Wilson and Nisbett, 1978).

We located two studies in which participants were asked specifically about their own behavior in relation to RCR (Davis et al., In press; Martinson et al., 2006). Martinson et al. (2006) were among the first to develop a self-report measure on research integrity. Their measure consists of a list of 33 problematic research behaviors that range from innocuous research behaviors, to behaviors that would be more questionable, to RM. Researchers who received NIH support in the past were asked to respond to the items, indicating that they either had (a) engaged in the behavior or (b) not engaged in the behavior, thus creating a yes/no response to QRPs and RM. This measure is a starting point for gauging the actual prevalence of RM and QRPs. Researchers, however, have shown mixed results on the validity of self-reported, socially-disapproved behavior (e.g., Del Boca and Noll, 2000; Golub et al., 2002). Due to the difficulty in measuring RM and QRPs (Anderson, 1996), and the fact that most researchers report on others’ behaviors instead of their own (Wiles et al., 2006), asking participants a yes/no response may result in more extreme underreporting of actual research behaviors. Nevertheless, Martinson and colleagues did find percentages that ranged from 0.2% to over 20% on problematic research behaviors.

Another study in which participants self-reported their own behavior was that of Davis et al. (In press). Respondents were asked to indicate how likely they were to engage in the same behavior as the hypothetical researcher within a vignette. Davis et al. found 2% to 23% of the respondents indicated they would be likely to engage in behaviors that departed from RCR. This study employed eight vignettes grounded in the American Counseling Association’s ethical codes regarding research. Thus, this measure would not be broad enough to assess for a variety of RCR, specifically the nine RCR areas specified by ORI (ORI, 2000), and would not be applicable to academic research areas outside the counseling field.

The lack of an empirical measure to assess self-reported QRPs and RM that is inclusive of a wide variety of questionable research behaviors is problematic. Thus, the purpose of the current study was to develop a self-report measure to assess a range of QRPs and RM.

Methodology

Sample

The population used for the development of the RCRM consisted of faculty members within five disciplines (counseling, criminology and criminal justice, economics, psychology, and special education) at research intensive/extensive private universities as defined by Carnegie Foundation for the Advancement of Teaching. The five specific disciplines were selected from the social science realm because few studies have focused on this group of researchers. These five were selected based on the existence or absence of a larger overarching professional ethical code and research ethics within the ethical code, each determined by the authors' independent examination of websites of the national professional organizations of various social science disciplines. Ethical codes were a decisive factor in the selection of disciplines due to the discernment that one of the main foundations of an environment that promotes research integrity is the existence of a research mission and expected standards of ethical research behavior (Recchia Jeffers, 2005). Therefore, we wanted to include disciplines that both had and did not have ethical codes addressing research.

Of the five disciplines selected, two had ethical codes that included research ethics incorporating the majority of the nine RCR areas stated by ORI (i.e., counseling: American Counseling Association, 2005; psychology: American Psychological Association, 2002); one discipline had an ethical code, but did not include research ethics (i.e., special education: Council for Exceptional Children, 2005); and two disciplines lacked overarching professional ethical codes (i.e., criminology/criminal justice: American Society of Criminology; economics: American Economic Association). This created a pool of 87 research intensive/extensive private universities which had at least one of five disciplines listed as a department, yielding a total of 3,476 possible participants. Some of the selected disciplines were similar (e.g., counseling and psychology) while others were more distinct (e.g., criminology/ criminal justice and special education). These choices were deliberate so it could be determined if the RCRM was applicable across different disciplines.

The nine RCR areas discussed above were created by ORI (2000) as areas which should be incorporated into instruction of research. These areas include: (a) data acquisition and management, (b) mentor/trainee responsibilities, (c) publication practices and responsible authorship, (d) peer review, (e) collaborative science, (f) human subjects, (g) research involving animals, (h) RM, and (i) conflict of interest.

Overall Procedure

We conducted three waves of data collection to pilot test the RCRM. The first wave was conducted solely to ensure the readability of the items and to determine which needed to be edited, deleted, or added. The purpose of the second wave was to assess factor structure and make any final edits. The third and final wave of the pilot testing was administered to assess reliability and factor structure, as well as to test validity against a social desirability measure. All waves were conducted using a Web-based survey. Each wave was approved by The University of North Carolina at Greensboro's Institutional Review Board, and each participant was provided with information about the voluntary nature of the study and the anonymity of their responses.

In each pilot wave of data collection, different sample sizes were used, each corresponding with adequate sample size for factor analysis, as well as for the specified purpose of the wave. For example, power was not considered in selecting wave 1 sample size due to the purpose of determining readability as well as the items to remain on the RCRM, but it was a consideration in waves 2 and 3.

Determining an adequate sample size for factor analysis is based on many aspects of the data. Research has shown that fairly small samples (less than 100 under specific criteria) and between 100 and 200 if several items are used to define each factor (MacCallum et al., 1999). In addition, other researchers indicate the number of participants included should more than the number of items included on the instrument (Aleamoni, 1976), that the sample size should be no less than 50 participants with 100 being adequate in some situations (Sapnas and Zeller, 2002), and that as long as there are more than 10 variables with moderate structure coefficients representing a construct on the instrument, a sample size of 150 should be adequate to obtain an accurate

solution (Guadagnoli and Velicer, 1988). All of these recommendations were taken into consideration for waves 2 and 3.

Pilot Study Wave 1

The Initial RCRM

The conceptual basis of the RCRM included eight of the nine RCR areas (ORI, 2000); research with animals was not included in the development of the measure due to the lack of relevance across all disciplines. The draft RCRM contained a pool of 96 items across the 8 remaining RCR areas. Item format of the RCRM was determined based on documented limitations of self-reported, socially disapproved behavior (e.g., Del Boca and Noll, 2000; Golub et al., 2002), difficulty of measuring QRPs and RM (Anderson, 1996), and participants' tendency to report on others' behaviors instead of their own (Wiles et al., 2006). The RCRM was designed so that participants would indicate "how likely" they would be to engage in the specified research behavior rather than "have they" engaged in the behavior. This method was used previously with success (see Davis et al., In press). Participants were given a 6-point Likert-type scale (1 = extremely unlikely to engage in the behavior, 6=extremely likely). To guard against response bias, as discussed by Converse and Presser (1986), approximately half of the items were reverse coded.

Procedure

The first pilot study was designed to assess for readability of the original 96 items of the RCRM, as well as to examine which items contained the most variance and correlated with other items so they could be retained in the RCRM. The sample consisted of two universities randomly selected from the original 87. Respondents (n=103) received two e-mails. The first provided an introduction to the study and a URL link to the consent form and survey page. The second email provided the same information and was sent two weeks later. The survey was distributed via email from Survey Monkey. Survey Monkey has the ability to house participants' e-mail addresses and track who has responded and who has not, so that those that have responded do not receive future e-mails about the study. However, we requested Survey Monkey to not collect respondents' IP addresses and e-mails from the database to ensure anonymity. A total of 30 faculty responded from both universities (29% response rate), with 14 to 25 valid responses depending upon the scale.

Results

Items were analyzed based on classical test theory (CTT) item statistics. Because the sample size involved was extremely small (n=14 to 25), item analyses were not used with absolute rules for dropping or retaining items. Rather, within each of the 8 RCR areas, items that exhibited low variance or small item-total correlations were flagged for further review. Two of the items thus flagged were retained because their content was central to the construct of interest, and respondents did not indicate a problem with readability. That is, a desire to maintain the integrity of our construct definitions led us to retain some items that had been flagged for empirical reasons. We did not want psychometrics solely to dictate our construct definitions (McDonald, 1999) with the knowledge that a statistically ideal assessment that omits items central to the construct will not provide valid information (Allen and Yen, 1979).

Some items were revised or removed when the review suggested the item would continue to exhibit unsatisfactory performance on the scale. Most items removed at this stage addressed research issues with likely socially desirable responses (e.g., Recommend rejection of a publishable manuscript submitted for publication that you believe should have cited your work, but the author did not). We assumed that the low item variance associated with these items was due to wording of the question (i.e., confusing item or written in a more extreme behavioral format), and that, regardless of the domain appropriateness, responses would continue to be endorsed strongly in a socially desirable direction. Seventy-five items remained after this initial scale revision. Based on feedback and written responses to some items (e.g., "I never experienced this behavior before;" "I do not collect this type of data, thus this item is NA"), the instructions were changed to ask participants to respond to the items as hypothetical situations.

Pilot Study Wave 2

Revisions to the RCRM

The revised RCRM contained 75 items representing 8 areas of RCR. The instructions provided to respondents were as follows: “Listed below are multiple hypothetical situations or behaviors that could occur in research or mentoring. Some of the situations you may have engaged in or experienced, while others you may not have encountered. Please take a moment to consider how you would respond to each of the hypothetical situations/behaviors below. How likely would you be to....” Participants in this iteration also were asked to respond using the same 6-point Likert-type scale that was used in Wave 1. Participants were asked if they had any comments on the survey or the items.

Procedure

For all pilot tests, the sampling frame consisted of 87 private research intensive/extensive universities (as defined by Carnegie Foundation for the Advancement of Teaching) that contained at least one of the five disciplines. Knowing that the response rate in pilot wave 1 was 29%, the researchers wanted to ensure that the responding sample size would be large enough to begin exploratory factor analysis in pilot wave 2 in case of a low response rate. Therefore, for the second wave of the pilot test, 21 of the remaining 85 universities (excluding the two used in the first wave) were selected randomly.

This resulted in a total sampling frame of 936 participants, well above the recommended sample sizes for factor analysis (Aleamoni, 1976; Guadagnoli and Velicer, 1988; MacCallum et al., 1999). The same method of distributing two emails in wave 1 was used in wave 2. In addition to the previous method, participants in wave 2 were given the possibility of entering themselves into a raffle to receive \$75. Almost 15% of participants responded to the survey (n = 136). This low response rate may have occurred for a number of reasons, including the following: (a) the survey was submitted during the middle of the spring semester, a time when many faculty are involved in midterm exams; (b) the information being collected in the survey is sensitive in nature and some may have feared the possibility of non-anonymity on a Web-based survey; (c) some of the e-mails were undeliverable; and (d) the survey was lengthy (75 items on the RCRM in addition to multiple demographics items). In a question asked at the end of the survey, no participants suggested any items were unreadable. In addition, based on new instructions for the RCRM, no respondent indicated that any item was not applicable to his or her research activities.

Of the 136 responses, 83 participants responded to every item on the survey. These individuals’ responses were used for the analyses in this second wave. While this response rate resulted in a low sample size, the sample size of 136 (83 with complete data) is still sufficient to run beginning exploratory analyses (Aleamoni, 1976; Sapanas and Zeller, 2002). However, due to the sample size being too small for reliable results from a principle components analysis (PCA), and consistent with our prior procedures in Wave 1, a combination of CTT and PCA with promax rotation was used to guide development process. Promax rotation is almost always a good choice when there is reason to believe that factors might be correlated (Thompson, 2004). Because the most coherent interpretation of the PCA and scree plots was based on a single-factor model, scale revisions at this stage focused on developing a coherent overall measure. As a result of the PCA findings and the CTT item-level statistics, a final set of items was chosen that retained a domain balance across the eight core areas for RCR and displayed reasonable item functioning. The revised form consisted of 42 items with an alpha of .89. No participants from Wave 2 indicated difficulty in readability or clarity of any of the items or instructions; therefore, no further edits of the 42 items or the instructions were made.

Pilot Study Wave 3

Final Implementation of the RCRM and Other Instruments

The final version of the RCRM included 42 behavioral items relating to research activities that are based on 8 of the 9 ORI RCR areas (with 4 to 8 items representing each RCR area). Participants were asked to respond to each item with the likelihood that they would engage in the behavior. Responses were recorded on the same 6-point Likert-type scale that was used in Waves 1 and 2 (1 = extremely unlikely; 6 = extremely likely). RCRM scores can range from 42 to 252, with higher scores indicating a greater likelihood of engaging in RCR and

lower scores representing the likelihood of engaging in QRPs or RM. Nineteen of the 42 items are reverse coded to address the possibility of response bias (Converse and Presser, 1986) .

In addition to the RCRM, participants were asked to complete a social desirability scale. Crowne and Marlow (1964) presented the Personal Reaction Inventory (PRI) as a measure of a person's need for approval. One of the main uses of the Marlow–Crowne PRI is to determine whether respondents answer in a socially desirable manner on a scale of interest, which is indicated by a moderate to high correlation between the two scales. A short form of the Marlow–Crowne social desirability scale (Ray, 1984) was used to assess for social desirability in the RCRM. The 8 items have alpha reliabilities of .74 to .77 (Ray, 1984). Respondents are asked to indicate yes or no to each of the eight items, but they can also indicate “not sure.” Three of the 8 items are reverse coded. Higher scores indicate less likelihood to act in a socially desirable manner.

Finally, each participant was asked to provide demographic variables, including sex, age, academic faculty rank/position, degree, departmental discipline, overarching professional organization, productivity (measured through number of publications, presentations, and current positions/committees held), and the number of grants. This information was collected to determine how involved the sample was in research.

Procedure

The final wave in the development of the RCRM consisted of the 1,867 individuals listed as faculty members within the five selected disciplines at the private research intensive/extensive universities within the population not used in waves 1 and 2. The same survey distribution method used in waves 1 and 2 was also used in this wave. Upon completion of the survey, respondents were given the opportunity to follow a different URL link, unrelated to the survey, to enter their email address to enter a raffle for a \$50 money order.

A total of 213 individuals responded to the survey (response rate 11.5%). Again, while this response rate was low, it remains an adequate sample size to run exploratory analysis of items deemed to represent a construct of RCR (Aleamoni, 1976, MacCallum et al., 1999; Sappas and Zeller, 2002) and is larger than the sample of 150 needed to obtain an accurate solution (Guadagnoli and Velicer, 1988). Of these 213 respondents, 55% were female and 86% identified as Caucasian, with an average age of 47.60 years old ($SD=11.13$). The primary discipline responding was special education (63.8%), with 11.2% respondents from economics, 5.6% from psychology, .9% from counseling, .9% from criminology/criminal justice, and 17.4% indicating “other” as their departmental discipline (e.g., general education, political science, sociology or social work).

A majority of the respondents were in tenure-track positions (89%), including 26% assistant professors, 35% associate, and 28% full. Remaining respondents included 2.3% instructors, 3.3% adjunct professors, 1.4% emeritus, and 4.2% other. Respondents had been working as faculty members for an average of 14.13 years ($SD = 11.13$), and had been conducting research for an average of 18.22 years ($SD = 13.26$).

Respondents reported an average of 7.42 publications ($SD = 6.07$), and an average of 7.41 presentations in the past two years ($SD = 7.85$). Most (72%) indicated they mentored students, and 75% indicated they had received a grant recently or currently were working on a grant. The dollar amount of the grants ranged from \$1,500 to \$14,000,000, with an average of 1.61 ($SD = 1.26$) recent grants per respondent.

Results

The dimensionality of the 42-item measure was examined using principal component analysis (PCA); for more information on this methodology, and the subsequent graphs and analyses discussed here, see Thompson (2004) or other factor analysis resources. Of the 213 respondents, 153 participants responded to every item on the survey. These individuals' responses were used for the analyses of the third wave. This sample size is somewhat low for an application of PCA with 42 observed variables, so results were interpreted with caution. Examination of eigenvalues suggested that using more than one factor to explain the measure should be explored further. Using an eigenvalue greater than 1 or using the scree method suggested retaining several factors (as many as 14). Since retaining 14 factors would not be parsimonious, several possible factor structures were explored.

With each multiple factor solution, promax rotation was used to increase interpretability. With each attempt (2, 3, and 4 factor solutions) content of items included in different factors did not lend themselves to easy interpretation. That is, items grouped within factors did not seem to share common themes or content.

This absence of a consistent factor structure can be seen when the first principal components are plotted against one another (Figure 1). In such a plot, items sharing a common factor will tend to be grouped together. If no pattern of item groups is apparent from visual inspection, there is no reason to conclude that there are separate factors influencing the item responses and thus, promax rotation becomes irrelevant. There are some distinct groupings of items. Notice that within each grouping, items from each RCR area often appear in more than one grouping. This mix of items across the general groupings is what prevents consistent interpretation of multiple factors. There are two exceptions to this interpretation. Using visual inspection of the principal component plot, two sets of items stand out.

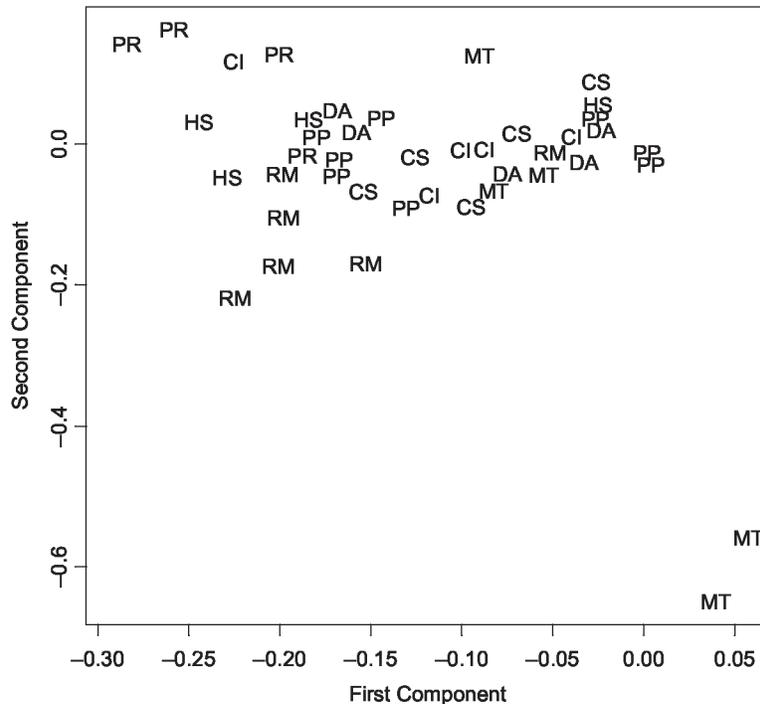


FIGURE 1 *First two principal components for Pilot Test Wave 3. Items from RCR areas are represented by two letters: DA = data acquisition, management, sharing, and ownership; MT = mentor/trainee relationships; PP = publication practices and responsible authorship; PR = peer review; CS = collaborative science; HS = human subjects; RM = research misconduct; and CI = conflict of interest and commitment.*

In the bottom right of Figure 1, two items from the mentor/ trainee relationships (MT) RCR area are clearly separated from the other items. These items share very similar content. They address whether authorship or acknowledgement is given to an assistant/student who assisted with data collection or data analysis. These items were endorsed with an approximate even distribution across all categories on the Likert scale much more so than other items. Taking their similarity and general response characteristics into account, there is little reason to interpret these items as belonging to a separate scale.

In the middle-left of Figure 1 there is an apparent grouping of RM items. These four items taken together yield a reliability of .85. This high level of internal consistency suggests that this area of RCR may support an independent scale, which makes sense since RM behaviors tend to be more aberrant in nature compared to QRPs. However, we do not suggest interpreting this set of items as a subscale at this time, as this four item subscale cannot be expected to cover the full domain of RM.

Based on these results, and an alpha of .84, a one-factor, unrotated solution seemed to be the best solution for the RCRM, suggesting construct validity. Therefore, the RCRM can be examined using a total sum score of all

42 items to assess the likelihood of engaging in RCR (score range: 42 to 252) or examined by focusing on individual items. Higher scores on the RCRM indicate a respondent self-reported a greater likelihood of engaging in RCR, while lower scores indicate the likelihood that one will engage in QRPs or RM.

The RCRM had a nonsignificant relationship with social desirability ($r(151) = .037, p = .65$), suggesting that participants did not respond to the RCRM in a socially desirable manner. Finally, due to examining the RCRM across five diverse disciplines, differences among disciplines was tested. An ANOVA indicated no significant difference in responses on the RCRM were found across disciplines with sufficient numbers of responses ($n > 10$; psychology, economics, special education, and other) ($F(3, 150) = .354, p > .10$).

The overall mean score on the RCRM was 214.10 ($SD = 15.80$; range 163 to 243). Since the maximum range of the RCRM is 252, this suggests that the majority of faculty in the current sample reported the likelihood of engaging in RCR, with only a few likely to engage in some QRPs (see lower range). However, while the scale results are provided for informational purposes, they should be taken with caution and not generalized until further reliability and validity analyses are done on the RCRM.

Discussion

The main purpose of this study was to develop an instrument that would measure researchers' self-reported RCR behaviors. We found that 42 items loaded on one factor of responsible conduct of research—with higher scores indicating a greater likelihood to engage in RCR and lower scores indicating a greater likelihood to engage in QRPs or RM.

The results of this study are not generalizable to all academic researchers. We obtained a low response rate in the pilot tests, are unclear of the differences that might exist between respondents and non-respondents, and limited our sample to five professional disciplines and to private research intensive/extensive universities. However, the goal of this study was not to generalize, but instead to create a measure that assesses RCR. Accordingly, our results suggest the RCRM is one of the first RCR measures that is a self-reported, reliable ($\alpha = .84$ to $.89$) instrument. The RCRM shows some promise of being a valid assessment of QRP across multiple, diverse disciplines, in that the respondents reported in similar fashions with no significant differences across disciplines, and a one-factor solution was indicated. Despite the fact that self-reports of QRPs are sensitive in nature, the RCRM was found to have a low, nonsignificant correlation with social desirability, which strengthens divergent validity claims about the measure.

While this study shows promise for the RCRM, more studies do need to be conducted in order to gather further reliability and validity data within the same, as well as different, samples and disciplines (e.g., health sciences) to ensure the broader applicability of the RCRM. In addition to testing the RCRM on additional populations, other validity studies need to be conducted to assess divergent validity. While no other RCR measure exists to do a direct comparison, possible comparisons include correlating RCRM with moral reasoning or honesty and integrity measures.

After further validity and reliability analyses have been conducted and found, ultimately, the RCRM can be used to explore characteristics that cause departure from RCR. Such studies would inform the area of research ethics and integrity, possibly eliminating or minimizing engagement in QRPs and RM. Certainly, further research needs to be conducted on the psychometrics of the measure; however, the RCRM shows promise as a scaled measure of QRPs and RM.

References

- Alberts, B., and Shine, K. (1994). Scientists and the integrity of research, *Science*, 266: 1660–1661.
- Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques, *Educational and Psychological Measurement*, 36: 879–883.
- Allen, M. J., and Yen, W. M. (1979). *Introduction to Measurement Theory*. Lon Grove, Illinois: Waveland Press, Inc.

- American Counseling Association (2005). *ACA code of Ethics 2005*. Available at <http://www.counseling.org/Resources/CodeOfEthics/TP/Home/CT2.aspx>, last accessed on 18 March, 2008.
- American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Available at <http://www.apa.org/ethics/code2002.html>, last accessed 17 January, 2007.
- Anderson, M. S. (1996). Misconduct and departmental context: Evidence from the Acadia Institute's graduate education project, *Journal of Information Ethics*, 5: 15–33.
- Anderson, M. S., Seashore Louis, K., and Earle, J. (1994). Disciplinary and departmental effects on observations of faculty and graduate student misconduct, *The Journal of Higher Education*, 65: 331–350.
- Broome, M. E., Pryor, E., Habermann, B., Pulley, L., and Kincaid, H. (2006). The scientific misconduct questionnaire-revised (SMQ-R): Validation and psychometric testing, *Accountability in Research*, 12: 263–280.
- Converse, J. M., and Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–063. Beverly Hills, CA: Sage Publications.
- Council for Exceptional Children. (2005). *Professional standards*. Available at <http://www.idea practices.org/ps/>, last accessed 17 January, 2007.
- Crowne, D. P., and Marlowe, D. (1964). *The Approval Motive: Studies in Evaluative Dependence*. New York: Wiley.
- Davis, M. S., Riske, M. L., and Seaman, M., (2001). *Research Misconduct: An Inquiry into Etiology and Stigma*. Final Report Presented to the Office of Research Integrity. Amherst, OH: Justice Research and Advocacy, Inc.
- Davis, M. S., Wester, K. L., and King, B. (2008). Narcissism, entitlement, and questionable practices of research in counseling. A pilot study, *Journal of Counseling and Development*.
- de Vries, R., Anderson, M. S., and Martinson, B. C. (2006). Normal misbehavior: Scientists talk about the ethics of research, *Journal of Empirical Research on Human Research Ethics*, 1: 43–50.
- Del Boca, F. K., and Noll, J. A. (2000). Truth or consequences: The validity of self-report data in health services research on addictions, *Addiction*, 95: S347–S360.
- Gardner, W. (2006, December). *The social cost of scientific misconduct and questionable research practices*. Presentation given at the ORI Research Conference, Tampa, FL.
- Golub, A., Johnson B. D., Taylor, A., and Liberty, H. J. (2002). The validity of arrestees' self-reports: Variations across questions and persons, *Justice Quarterly*, 19: 477–502.
- Guadagnoli, E., and Velicer, W. F. (1988). Relation of sample size to the stability of component patterns, *Psychological Bulletin*, 103: 265–275.
- Institute of Medicine National Research Council. (2002). *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct*. Washington, D.C.: The National Academics Press.
- McDonald, R. P. (1999) *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Martinson, B. C., Anderson, M. S., Crain, A. L., and de Vries, R. (2006). Scientists' perceptions of organizational justice and self-reported misbehaviors, *Journal of Empirical Research on Human Research Ethics*, 1: 51–66.
- MacCallum, R. C., Widaman, K. F., Zhang, S., and Hong, S. (1999). Sample size in factor Analysis, *Psychological Methods*, 4: 84–99.
- Office of Research Integrity. (2000). *PHS policy on instruction in the responsible conduct of research (RCR) – suspended*. Available at <http://ori.dhhs.gov/html/programs/finalpolicy.asp>, last accessed 12 March, 2004.
- Office of Research Integrity. (2003). Research on research integrity, Retrieved online February 20, 2004, at <http://grants1.nih.gov/grants/guide/rfa-files/RFA-NS-03-001.html>.
- Office of Research Integrity. (2006). Handling misconduct – case summaries. Available at: <http://ori.dhhs.gov/misconduct/cases/>, last accessed 20 February, 2004.
- Rankin, M., and Esteves, M. (1997). Perceptions of scientific misconduct in nursing, *Nursing Research*, 46: 270–276.
- Ray, J. (1984). The reliability of short social desirability scales, *Journal of Social Psychology*, 123: 133–134.
- Recchia Jeffers, B. (2005). Research environments that promote integrity, *Nursing Research*, 54: 63–70.
- Sapnas, K. G., and Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement, *Journal of Nursing Measurement*, 10: 135–154.

- Shafir, S., and Kennedy, D. (1998). Research misconduct: Media exaggerate results of a survey, *The Scientist*, 12: 0. Available at: http://www.the-scientist.com/yr1998/june/opin_980622.html, last accessed 15 July, 2007.
- Steneck, N. H. (2003). The role of professional societies in promoting integrity in Research, *American Journal of Health and Behavior*, 27: S239–S247.
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions, *Science and Engineering Ethics*, 12: 53–74.
- Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis*. Washington, DC: American Psychological Association.
- Wiles, R., Charles, V., Crow, G., and Heath, S. (2006). Researching researchers: Lessons for research ethics, *Qualitative Research*, 6: 283–299.
- Wilson, T. D., and Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior, *Social Psychology*, 41: 118–131.