Measuring how people view biomedical research: Reliability and validity analysis of the Research Attitudes Questionnaire

Jonathan D. Rubright,
University of Delaware

Mark S. Cary,
University of Pennsylvania

Jason H. Karlawish, and
University of Pennsylvania

Scott Y. H. Kim
University of Michigan

Abstract

With increasing numbers of studies on research ethics and a need to improve the recruitment of research subjects, the ability to measure attitudes toward biomedical research has become important. The Research Attitudes Questionnaire is a significant predictor of the public's attitudes toward and willingness to participate in research, yet limited data are available on its psychometric properties. This study establishes the scale's internal consistency and dimensionality using a large Internet-based sample from the United States. One item was removed due to a poor item-total correlation, and three additional items were removed which formed a reverse-wording measurement artifact factor. With improved internal consistency and dimensionality, the seven-item version has the advantages of shorter administration time and improved psychometric properties.

Keywords

research participation; research attitudes; research ethics; exploratory factor analysis; confirmatory factor analysis; psychometrics

---

Research that requires the participation of volunteers presents a number of challenges. These include issues of informed consent, minimizing risk, and fair subject selection. Other challenges are more pragmatic and largely center upon efficiently recruiting and retaining eligible subjects.

Considerable scholarship has attempted to define the factors that predict whether someone is willing to participate in research. To date, studies suggest that core factors are attitudes about trust, benefit, and altruism (Karlawish et al., 2001; Sugarman et al., 1998). As valuable as this research is, however, a gap remains in the science: there is no scale that can predict willingness to participate.

Such a scale presents a number of potential advantages. It can help to classify persons who are likely to enroll, explain why studies fail to enroll in a timely manner, and potentially

---

Address correspondence to: Jonathan Rubright, 106 Pearson Hall, University of Delaware, Newark, DE 19716., rubright@udel.edu.

identify subjects likely to remain in a study. Also, as research ethics controversies arise, there is a need to incorporate public viewpoints into policies. A brief instrument that measures people's general attitudes toward biomedical research can help explain differing positions in such controversies.

The Research Attitudes Questionnaire (RAQ) is a recently developed scale designed to achieve these goals. By measuring how favorably or unfavorably one views biomedical research in general (Kim et al., 2005), the measure has been shown to be a significant predictor in studies of consent and willingness to participate in biomedical research studies, supporting predictive, construct, and criterion validity (Karlawish et al., 2009; Muroff, Hoerauf, & Kim, 2006).

However, limited data are available on its internal reliability and no data are available on factorial validity. Such analyses of psychometric properties are crucial to assure the scale is measuring a coherent construct with limited error. This study adds to the existing literature on the RAQ by publishing its psychometric properties and exploring whether a shorter scale can be utilized.

Method

The RAQ Explained

The RAQ is an eleven-item scale with five Likert options, strongly disagree to strongly agree. A final score is produced by summing all individual item scores, with final scores ranging from 11 to 55 and higher scores indicating a higher value placed on research. To achieve this directionality, questions 2, 5, 7, and 10 are reverse coded. All questions are shown in Table 1.

Data

Data were taken from two samples. The first sample is from an Internet study of the general public on perceptions of stigma associated with research involving persons with mental illness. For this sample, subjects responded to e-mails requesting participation. Subjects were randomly selected from a survey panel maintained by a survey sampling firm, with oversampling of minorities and older adults to assure balanced participation. The sample is 80.6% White, 54.7% female, and average 48.1 (SD = 14.7) years of age. Greater detail on the sample is provided elsewhere (Muroff et al., 2006). This sample, referred to here as the STIGMA sample, was listwise deleted (n = 2,406, 76 cases deleted) so that all subjects had complete data on the RAQ. The second sample (n = 555, 3 cases listwise deleted due to missing data on the RAQ) is from a one-hour-long face-to-face interview on research advance planning with persons over the age of 65 in the Philadelphia, Pennsylvania area (Karlawish et al., 2009). This sample, referred to here as the RAP sample, is 61.2% White, 58.7% female, and average 76.8 (SD = 6.7) years of age.

Organizing Framework for Analysis

To improve the scale's psychometric properties, our analyses followed a standard process: an initial item analysis and reliability study to identify items that are poorly related to the rest of the scale, an exploratory factor analysis (EFA) to search for underlying factor structure, and two confirmatory factor analyses (CFAs) to determine the fit of the structures found in the EFA. Along the way, items which do not adhere to the theoretical expectations of a unidimensional scale will be considered for removal so that the final scale to be tested in the confirmatory models has the best possible psychometric properties. Additionally, we consider practicality. We aim for a short, easy-to-score scale that can be straightforwardly

employed by researchers, in a variety of contexts, and within any statistical framework. Additional detail on these analyses is provided below.

A key consideration in psychometric analyses is that of cross-validation, "a rechecking of the statistical properties of the test based on a new sample" (Allen & Yen, 1979, p. 139) so that data used for exploratory analyses are not used for confirmatory analyses. In order to adhere to this recommendation, the STIGMA sample was randomly split in half. The first half (n = 1201) was used for the initial reliability study and exploratory factor analysis. The second half (n = 1205) was used for a confirmatory factor analysis replication for stronger structural evidence to emerge (Goldberg & Velicer, 2006). This change in data was done to ensure that the same data used to generate plausible factor structures were not used to test these relationships in the confirmatory factor models. Finally, the RAP sample is used to cross-validate the confirmatory models established from the STIGMA sample.

Reliability

Item-total correlations, along with Cronbach's coefficient alpha, assessed the preliminary internal reliability of the RAQ (Cronbach, 1951).

Exploratory Factor Analysis

Exploratory factor analysis (EFA) identified integral constructs because of the uncertainty surrounding the underlying structure of the RAQ (Browne, 2001). Principal axis factor analysis was employed given its relative tolerance of multivariate nonnormality and its superior recovery of weak factors (Briggs & MacCallum, 2003; Cudeck, 2000; Fabrigar et al., 1999). Communalities were estimated through squared multiple correlations and iterated to produce final estimates (Gorsuch, 2003). Because it was assumed that factors would be correlated, a Promax rotation was utilized with $k = 3$ (Tataryn, Wood, & Gorsuch, 1999).

Implementation of solitary criteria to determine the correct number of factors to retain and rotate tends to either under- or overestimate the number of true latent dimensions (Gorsuch, 1983; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). Accordingly, each model was evaluated against the following five rules: (1) eigenvalues greater than 1.0 (Kaiser, 1960), (2) scree (Cattell, 1966), (3) Glorfeld's (Glorfeld, 1995) extension of parallel analysis (PA) (Horn, 1965), (4) minimum average parcels (MAP) (Velicer, 1976), and (5) interpretability (Fabrigar et al., 1999; Gorsuch, 1983).

Confirmatory Factor Analysis

After determining a plausible factor structure for these data, a confirmatory factor analysis (CFA) was conducted to determine model fit. Since the chi-square hypothesis test of the plausibility of hypothesized relationships is typically significant with large samples, we relied on the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the comparative fit index (CFI) as measures of model fit (Bentler, 1990; Browne & Cudeck, 1993). RMSEA relates model fit to degrees of freedom, with RMSEA < 0.10 acceptable fit. SRMR is the average difference between hypothesized and observed covariances in the model using standardized residuals, with similar interpretation as RMSEA. The CFI compares the measurement model to a null model, with convention values of acceptable fit > 0.90.

CFA models were constructed in LISREL 8.80 (Joreskog & Sorbom, 2009), with all other analyses in SAS 9.1 (SAS 9.1).

Results

Reliability Analysis

The initial alpha value with all eleven items on the first half of the STIGMA sample was 0.78. All items showed good item-total correlations with the exception of q7, with r = 0.03. Removal of any item, except q7, reduced the alpha level. Additionally, interviewers from two studies using the RAQ (Karlawish et al., 2008; Karlawish et al., 2009) report research subjects had the most difficulty answering this item (personal communication). For these reasons, q7 was dropped from the scale and from the following EFA. After removing q7, alpha increased to 0.81, well above the 0.70 criterion recommended by leading measurement textbooks (Allen & Yen, 1979; Thorndike, 1982).

Exploratory Factor Analysis

Using the first half of the STIGMA sample without q7, the Kaiser-Meyer-Olkin (KMO) (Kaiser, 1974) statistic was 0.87 and well above the 0.60 minimum suggested for sampling adequacy appropriate for factor analysis (Kline, 1994). Parallel analysis, scree, and Kaiser's criterion all suggested that two factors be retained, yet MAP indicated a one-factor solution, suggesting that factor analysis was unnecessary. The two-factor solution was rotated because most decision rules pointed to two factors, and the two-factor solution was interpretable.

Table 2 presents the rotated pattern matrix for the two-factor solution. The two factors were interpreted according to the magnitude and meaning of their salient pattern coefficients. Notably, Factor II has appreciable loadings from all of the negatively worded items. The correlation between these two retained factors is high ($r = 0.59$), indicating that the two dimensions are oblique and share a significant portion of common variance. Coefficient alpha was used to estimate internal-consistency reliability for both factors: 0.79 and 0.64 for Factor I and Factor II, respectively.

The loadings on the second factor, along with the significant relationship between both factors, suggest a measurement artifact of reverse wording. An exploratory factor analysis was performed with the reverse-scored items (rq2, rq5, and rq10) removed from the scale. The loadings on the remaining seven items are displayed in Table 3.

Confirmatory Factor Analysis

After determining a plausible factor structure for these data from the EFA, a CFA was conducted to determine model fit using the second half of the STIGMA sample. This change in data was done to ensure that the same data used to generate plausible factor structures were not used to test the confirmatory factor models. A total of four models were examined using these data. Model 1 tested whether the scale as it stands could be considered unidimensional, with all ten items loading on one factor. Model 2 used seven items loading on one factor and the three reverse-coded questions loading on a second correlated factor, as suggested by the EFA. Model 3 tested the model fit using only the positively worded seven items loading onto one factor, removing the three reverse-coded items. Model 4 used all ten items loading on one factor, with a second, uncorrelated method factor with loadings from the reverse-coded items.

All models show adequate SRMR statistics (SRMR <= 0.06). However, Model 1 showed borderline fit considering RMSEA (0.099) and CFI (0.91), suggesting that the original scale may not be considered unidimensional. In Model 2, allowing the reverse-coded items to load onto a separate factor as suggested by the EFA improved the model (RMSEA = 0.076, CFI = 0.95).

Model 3 considered what would happen if the reverse-coded items were dropped from the scale and the remaining seven items were allowed to load onto one factor that does show adequate internal consistency. This model had adequate CFI (0.95), yet borderline RMSEA fit (0.092). Model 4 had acceptable RMSEA (0.078) and CFI (0.95), and modeled the thinking from the EFA that the scale was essentially unidimensional but for the methods factor shared by items rq2, rq5, and rq10. The seven-item/one-factor model from Model 3 was tested using the RAP sample as well; this model showed adequate fit (RMSEA = 0.079, CFI = 0.94).

Discussion

Previous studies have identified the Research Attitudes Questionnaire as an important predictor in willingness to participate. However, limited data have been available on the properties of the scale. This study presents the psychometric properties of the RAQ. In particular, we present internal consistency, an exploratory factor analysis, and a confirmatory factor analysis. After removing one item which was holding down internal consistency and a factor structure was identified using an EFA, a series of four CFA models were run.

The fit statistics show that most models fit acceptably. However, Model 1 has a borderline fit, which is not surprising since the EFA suggests a two-factor solution. Model 2 fits considerably better. However, the second factor is not theoretically independent because it is significantly correlated with the first factor and is made up of items which only share a methodological component. Additionally, Factor II does not have adequate internal consistency to serve as an independently functioning sub-scale. Thus, the two-factor model suggested by the EFA may not be acceptable for use in practice.

Model 4 fits well, and makes explicit the measurement artifact of reverse coding. Yet, using this type of model with an uncorrelated method factor requires all analyses be performed in a Structural Equation Modeling framework. And, as this scale is often used in biostatistical settings where Structural Equation Modeling is not routinely utilized, Model 4 may not be easily used in practice. Overall, Model 3 is the most feasible for use in practice. Notably, the structure of Model 3, established from halves of the STIGMA sample, generalizes well to the RAP sample, supporting stability of structure across samples.

Best Practices

This study provides evidence of reliability and validity for a scale on research attitudes. For research to proceed on any specific construct, measuring the construct well and without error is often a good place for researchers to start. Prior to this study, this scale already had solid evidence of credibility: it was created by an expert in the field and correlated highly and in expected directions with hypothesized variables in empirical bioethics studies. Still, having a more refined scale that more closely aligns with the ideal of measuring a single latent construct without error gives more confidence to users who want to utilize this construct in future empirical and theoretical work.

We recommend the seven-item RAQ version for general usage. It can be scored using a simple additive scale that is quick to administer, easy to score, and can be used with a variety of modeling techniques that do not require co-varying out a methodological component. Additionally, the model structure generalizes across samples. The seven-item factor model chosen with the STIGMA sample fits well in the RAP sample. This is especially noteworthy due to the different nature of the two samples. For these reasons, this reduced form is recommended for future studies on attitudes toward research.

Research Agenda

In terms of the scale itself, future studies should test whether this scale shows measurement invariance across theoretically meaningful groups, be they groups stratified by age, gender, or racial identity. Measurement invariance is an important way of assessing whether the construct is construed similarly across members of distinct groups. Some areas, such as achievement testing in education, use invariance studies to identify items biased against certain groups for removal. In the context of research attitudes, items that act differently across distinct groups may reveal ways in which persons of different ages or backgrounds feel about various aspects of the research process.

With a psychometrically defensible scale on research attitudes in hand, a number of future studies are possible. First, categorizing individuals by their research attitudes may give focus for researchers during the planning phase of studies so that they can target individuals most likely to participate, thus reducing recruitment costs and possibly even the risk of drop-out. As previous research has correlated research attitudes with hypothetical research involvement, future studies can empirically test whether the scale predicts actual willingness to participate. Beyond this potential practical use, the RAQ may be an important tool for policy research, since how much the public values biomedical research is an important variable to take into account in biomedical research controversies and public policy dilemmas. The refined version of the RAQ presented here can speak to these topics.

Educational Implications

As further studies show the value of the RAQ-7 to identify persons more or less willing to participate, the scale can become part of the training of clinical researchers and research ethicists. As clinical investigators and research ethicists review the relevant studies, discuss the scale and its items, and use it in the field, they will begin to develop a common language to talk about why people do and do not want to be in research and the effects of interventions to improve attitudes. Ultimately, the field may discover clinically and ethically relevant cut-points on the scale that categorize a person's attitudes about research.

Acknowledgments

References

Allen, M.; Yen, W. Introduction to measurement theory. Monterey, CA: Brooks/Cole; 1979.

Bentler P. Comparative fit indexes in structural models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]

Briggs N, MacCallum R. Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. Multivariate Behavioral Research. 2003; 38(1):25–56.

Browne M. An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research. 2001; 36(1):111–150.

Browne, M.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-162.

Cattell R. The scree test for the number of factors. Multivariate Behavioral Research. 1966; 1(2):245–276.

Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3):297–334.

Cudeck, R. Exploratory factor analysis. In: Tinsley, H.; Brown, S., editors. Handbook of applied multivariate statistics and mathematical modeling. New York: Academic Press; 2000. p. 265-296.

Fabrigar L, Wegener D, MacCallum R, Strahan E. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. 1999; 4:272–299.

Glorfeld L. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement. 1995; 55:377–393.

Goldberg, L.; Velicer, W. Principles of exploratory factor analysis. In: Strack, S., editor. Differentiating normal and abnormal personality. New York: Springer; 2006. p. 209-337.

Gorsuch, R. Factor analysis. 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1983.

Gorsuch, R. Factor analysis. In: Shinka, J.; Velicer, F., editors. Handbook of psychology: Vol. 2, Research methods in psychology. Hoboken, NJ: John Wiley; 2003. p. 143-164.

Horn J. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965; 30:179–185. [PubMed: 14306381]

Joreskog, K.; Sorbom, D. LISREL (Version 8.80). Chicago, IL: Scientific Software International; 2009.

Kaiser H. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960; 20:141–151.

Kaiser H. An index of factorial simplicity. Psychometrika. 1974; 39:31–36.

Karlawish J, Cary M, Rubright J, TenHave T. How redesigning AD clinical trials might increase study partners' willingness to participate. Neurology. 2008; 71(23):1883–1888. [PubMed: 19047560]

Karlawish J, Casarett D, Klocinski J, Sankar P. How do AD patients and their caregivers decide whether to enroll in a clinical trial? Neurology. 2001; 56(6):789–792. [PubMed: 11274319]

Karlawish J, Rubright J, Casarett D, Cary M, TenHave T, Sankar P. Older adults' attitudes toward noncompetent subjects participating in Alzheimers research. American Journal of Psychiatry. 2009; 166(2):182–188. [PubMed: 18923066]

Kim S, Kim H, McCallum C, Tariot P. What do people at risk for Alzheimer disease think about surrogate consent for research? Neurology. 2005; 65(9):1395–1401. [PubMed: 16275826]

Kline, P. An easy guide to factor analysis. New York: Routledge; 1994.

Muroff J, Hoerauf S, Kim S. Is psychiatric research stigmatized? An experimental survey of the public. Schizophrenia Bulletin. 2006; 32(1):129–136. [PubMed: 16192411]

SAS (Version 9.1). Cary, NC: SAS Institute, Inc;

Sugarman J, Kass N, Goodman S, Perentesis P, Fernandes P, Faden R. What patients say about medical research. IRB: A Review of Human Subjects Research. 1998; 20(4):1–7.

Tataryn D, Wood J, Gorsuch R. Setting the value of k in promax: A Monte Carlo study. Educational and Psychological Measurement. 1999; 59:384–391.

Thorndike, R. Applied Psychometrics. Boston: Houghton Mifflin; 1982.

Velicer W. Determining the number of components from the matrix of partial correlations. Psychometrika. 1976; 41:321–327.

Velicer, W.; Eaton, C.; Fava, J. Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In: Goffin, R.; Helms, E., editors. Problems and solutions in human assessment: Honoring Douglas N Jackson at seventy. New York: Guilford; 2000. p. 41-71.

Zwick W, Velicer W. Comparison of five rules for determining the number of components to retain. Psychological Bulletin. 1986; 99:432–442.

Biographies

Jonathan D. Rubright is a doctoral student in the Research Methodology and Evaluation program at the University of Delaware. His work is in psychometrics, and he has administered the RAQ scale face-to-face to hundreds of research participants. He made substantial contributions to the study design, analysis and interpretation of data, and drafting of this article.

Mark S. Cary is a senior biostatistician at the University of Pennsylvania Biostatistics Analysis Center. He has over 30 years of experience in research design and data analysis,

and made substantial contributions to the study design, analysis and interpretation of data, and drafting of this article.

Jason H. Karlawish is Associate Professor of Medicine and Medical Ethics and Associate Director of the Alzheimer's Disease Center at the University of Pennsylvania. His research focuses on ethical issues in human subjects research. Recent studies have extensively explored why individuals do or do not choose to participate in research, and the RAQ has been an important measure collected as part of this work. Dr. Karlawish made substantial contributions to the study design, acquisition and interpretation of data, and drafting of this article.

Scott Y. H. Kim is Associate Professor of Psychiatry and Co-Director of Center for Bioethics and Social Sciences in Medicine at the University of Michigan. His expertise includes informed consent, decision-making capacity, and IRB issues. The original author of the RAQ, he made substantial contributions to the study design, acquisition, analysis, and interpretation of data, and drafting of this article.

TABLE 1

Research Attitude Questionnaire Items

Variable	Item
q1	I have a positive view about medical research in general.
rq2	Medical researchers are mainly motivated by personal gain.
q3	Medical researchers can be trusted to protect the interests of people who take part in their studies.
q4	We all have some responsibility to help others by volunteering for medical research.
rq5	Modern science does more harm than good.
q6	Society needs to devote more resources to medical research.
rq7	Medical research needs to be closely regulated in order to prevent harm to research participants.
q8	Participating in medical research is generally safe.
q9	If I volunteer for medical research, I know my personal information will be kept private and confidential.
rq10	A lot of emphasis on medical research and scientific progress is likely to harm research volunteers.
q11	Medical research will find cures for many major diseases during my lifetime.

Note: r = Reverse worded item.

TABLE 2

Exploratory Factor Analysis Rotated Pattern Matrix from Half of STIGMA Sample

Variable	Factor	
	I	II
q1	0.42	0.29
rq2	0.00	0.54
q3	0.66	0.06
q4	0.64	−0.07
rq5	0.00	0.62
q6	0.33	0.23
q8	0.62	0.02
q9	0.65	−0.01
rq10	0.06	0.52
q11	0.36	0.13

Note: r = Reverse worded item.

TABLE 3

Exploratory Factor Analysis Pattern Matrix from Half of STIGMA Sample, Items rq2, rq5, rq7, and rq10 Removed

Variable	Factor I
q1	0.63
q3	0.69
q4	0.59
q6	0.50
q8	0.62
q9	0.63
q11	0.45