The Research Identity Scale: Psychometric Analyses and Scale Refinement

Maribeth F. Jorgensen, William E. Schweinle

The 68-item Research Identity Scale (RIS) was informed through qualitative exploration of research identity development in master's-level counseling students and practitioners. Classical psychometric analyses revealed the items had strong validity and reliability and a single factor. A one-parameter Rasch analysis and item review was used to reduce the RIS to 21 items. The RIS offers counselor education programs the opportunity to promote and quantitatively assess research-related learning in counseling students.

Keywords: Research Identity Scale, research identity, research identity development, counselor education, counseling students

With increased accountability and training standards, professionals as well as professional training programs have to provide outcomes data (Gladding & Newsome, 2010). Traditionally, programs have assessed student learning through outcomes measures such as grade point averages, comprehensive exam scores, and state or national licensure exam scores. Because of the goals of various learning processes, it may be important to consider how to measure learning in different ways (e.g., change in behavior, attitude, identity) and specific to the various dimensions of professional counselor identity (e.g., researcher, advocate, supervisor, consultant). Previous research has focused on understanding how measures of research self-efficacy (Phillips & Russell, 1994) and research interest (Kahn & Scott, 1997) allow for an objective assessment of research-related learning in psychology and social work programs. The present research adds to previous literature by offering information about the development and applications of the Research Identity Scale (RIS), which may provide counseling programs with another approach to measure student learning.

Student Learning Outcomes

When deciding how to measure the outcomes of student learning, it is important that programs start with defining the student learning they want to take place (Warden & Benshoff, 2012). Student learning outcomes focus on intellectual and emotional growth in students as a result of what takes place during their training program (Hernon & Dugan, 2004). Student learning outcomes are often guided by the accreditation standards of a particular professional field. Within the field of counselor education, the Council for Accreditation of Counseling & Related Educational Programs (CACREP) is the accrediting agency. CACREP promotes quality training by defining learning standards and requiring programs to provide evidence of their effectiveness in meeting those standards. In relation to research, the 2016 CACREP standards require research to be a part of professional counselor identity development at both the entry level (e.g., master's level) and doctoral level. The CACREP research standards emphasize the need for counselors-in-training to learn the following:

> The importance of research in advancing the counseling profession, including how to critique research to inform counseling practice; identification of evidence-based counseling practices; needs assessments; development of outcome measures for counseling programs; evaluation of counseling interventions and programs; qualitative quantitative, and mixed research methods; designs in research and program evaluation; statistical methods used in conducting research

Maribeth F. Jorgensen, NCC, is an assistant professor at the University of South Dakota. William E. Schweinle is an associate professor at the University of South Dakota. Correspondence can be addressed to Maribeth Jorgensen, 414 East Clark Street, Vermillion, SD 57069, maribeth.jorgensen@usd.edu.

and program evaluation; analysis and use of data in counseling; ethically and culturally relevant strategies for conducting, interpreting, and reporting results of research and/or program evaluation. (CACREP, 2016, p .14)

These CACREP standards not only suggest that counselor development needs to include curriculum that focuses on and integrates research, but also identify a possible need to have measurement tools that specifically assess research-related learning (growth).

Research Learning Outcomes Measures

The Self-Efficacy in Research Measure (SERM) was designed by Phillips and Russell (1994) to measure research self-efficacy, which is similar to the construct of research identity. The SERM is a 33-item scale with four subscales: practical research skills, quantitative and computer skills, research design skills, and writing skills. This scale is internally consistent ($\alpha = .96$) and scores highly correlate with other components such as research training environment and research productivity. The SERM has been adapted for assessment in psychology (Kahn & Scott, 1997) and social work programs (Holden, Barker, Meenaghan, & Rosenberg, 1999).

Similarly, the Research Self-Efficacy Scale (RSES) developed by Holden and colleagues (1999) uses aspects of the SERM (Phillips & Russell, 1994), but includes only nine items to measure changes in research self-efficacy as an outcome of research curriculum in a social work program. The scale has excellent internal consistency ($\alpha = .94$) and differences between pre- and post-tests were shown to be statistically significant. Investigators have noticed the value of this scale and have applied it to measure the effectiveness of research courses in social work training programs (Unrau & Beck, 2004; Unrau & Grinnell, 2005).

Unrau and Beck (2004) reported that social work students gained confidence in research when they received courses on research methodology. Students gained most from activities outside their research courses, such as participating in research with faculty members. Following up, Unrau and Grinnell (2005) administered the scale prior to the start of the semester and at the end of the semester to measure change in social work students' confidence in doing research tasks. Overall, social work students varied greatly in their confidence before taking research courses and made gains throughout the semester. Unrau and Grinnell stressed their results demonstrate the need for the use of pre- and post-tests to better gauge the way curriculum impacts how students experience research.

Previous literature supports the use of scales such as the SERM and RSES to measure the effectiveness of research-related curricula (Holden et al., 1999; Kahn & Scott, 1997; Unrau & Beck, 2004; Unrau & Grinnell, 2005). These findings also suggest the need to continue exploring the research dimension of professional identity. It seems particularly important to measure concepts such as research self-efficacy, research interest, and research productivity, all of which are a part of research identity (Jorgensen & Duncan, 2015a, 2015b).

Research Identity as a Learning Outcome

The concept of research identity (RI) has received minimal attention (Jorgensen & Duncan, 2015a, 2015b; Reisetter et al., 2004). Reisetter and colleagues (2004) described RI as a mental and emotional connection with research. Jorgensen and Duncan (2015a) described RI as the magnitude and quality of relationship with research; the allocation of research within a broader professional identity; and

a developmental process that occurs in stages. Scholars have focused on qualitatively exploring the construct of RI, which may give guidance around how to facilitate and examine RI at the program level (Jorgensen & Duncan, 2015a, 2015b; Reisetter et al., 2004). Also, the 2016 CACREP standards include language (e.g., knowledge of evidence-based practices, analysis and use of data in counseling) that favors curriculum that would promote RI. Although previous researchers have given the field prior knowledge of RI (Jorgensen & Duncan, 2015a, 2015b; Reisetter et al., 2004), there has been no focus on further exploring RI in a quantitative way and in the context of being a possible measure of student learning. The first author developed the RIS with the aim of assessing RI through a quantitative lens and augmenting traditional learning outcomes measures such as grades, grade point averages, and standardized test scores. There were three purposes for the current study: (a) to develop the RIS; (b) to examine the psychometric properties of the RIS from a classical testing approach; and (c) to refine the items through future analysis based on the item response theory (Nunnally & Bernstein, 1994). Two research questions guided this study: (a) What are the psychometric properties of the RIS from a classical testing approach? and (b) What items remain after the application of an item response analysis?

Method

Participants

The participants consisted of a convenience sample of 170 undergraduate college students at a Pacific Northwest university. Sampling undergraduate students is a common practice when initially testing scale psychometric properties and employing item response analysis (Embretson & Reise, 2000; Heppner, Wampold, Owen, Thompson, & Wang, 2016). The mean age of the sample was 23.1 years (SD = 6.16) with 49 males (29%), 118 females (69%), and 3 (2%) who did not report gender. The racial identity composition of the participants was mostly homogenous: 112 identified as White (not Hispanic); one identified as American Indian or Alaska Native; 10 identified as Asian; three identified as Black or African American; eight identified as multiracial; 21 identified as Hispanic; three identified as "other"; and seven preferred not to answer.

Instruments

There were three instruments used in this study: a demographic questionnaire, the RSES, and the RIS.

Demographics questionnaire. Participants were asked to complete a demographic sheet that included five questions about age, gender, major, race, and current level of education; these identifiers did not pose risk to confidentiality of the participants. All information was stored on the Qualtrics database, which was password protected and only accessible by the primary investigator.

The RSES. The RSES was developed by Holden et al. (1999) to measure effectiveness of research education in social work training programs. The RSES has nine items that assess respondents' level of confidence with various research activities. The items are answered on a 0–100 scale with 0 indicating *cannot do at all*, 50 indicating *moderately certain I can do*, and 100 indicating *certainly can do*. The internal consistency of the scale is .94 at both pre- and post-measures. Holden and colleagues reported using an effect size estimate to assess construct validity but did not report these estimates, so there should be caution when assuming this form of validity.

RIS. The initial phase of this research involved the first author developing the 68 items on the RIS (contact first author for access) based on data from her qualitative work about research identity (Jorgensen & Duncan, 2015a). The themes from her qualitative research informed the development of items on the scale (Jorgensen & Duncan, 2015a). Rowan and Wulff (2007) have suggested that using

qualitative methods to inform scale development is appropriate, sufficient, and promotes high quality instrument construction.

The first step in developing the RIS items involved the first author analyzing the themes that surfaced during interviews with participants in her qualitative work. This process helped inform the items that could be used to quantitatively measure RI. For example, one theme was Internal Facilitators. Jorgensen and Duncan (2015a) reported that, "participants explained the code of internal facilitators as self-motivation, time management, research self-efficacy, innate traits and thinking styles, interest, curiosity, enjoyment in the research process, willingness to take risks, being open-minded, and future goals" (p. 24). An example of scale items that were operationalized from the theme Internal Facilitators included: 1) I am internally motivated to be involved with research on some level; 2) I am willing to take risks around research; 3) Research will help me meet future goals; and 4) I am a reflective thinker. The first author used that same process when operationalizing each of the qualitative themes into items on the RIS. There were eight themes of RI development (Jorgensen & Duncan, 2015a). Overall, the number of items per theme was proportionate to the strength of theme, as determined by how often it was coded in the qualitative data. After the scale was developed, the second author reviewed the scale items and cross-checked items with the themes and subthemes from the qualitative studies to evaluate face validity (Nunnally & Bernstein, 1994).

The items on the RIS are short with easily understandable terms in order to avoid misunderstanding and reduce perceived cost of responding (Dillman, Smyth, & Christian, 2009). According to the Flesch Reading Ease calculator, the reading level of the scale is 7th grade (Readability Test Tool, n.d.). The format of answers to each item is forced choice. According to Dillman et al. (2009), a forced-choice format "lets the respondent focus memory and cognitive processing efforts on one option at a time" (p. 130). Individuals completing the scale are asked to read each question or phrase and respond either *yes* or *no*. To score the scale, a *yes* would be scored as one and a *no* would be scored as zero. Eighteen items are reverse-scored (item numbers 11, 23, 28, 32, 39, 41, 42, 43, 45, 48, 51, 53, 54, 58, 59, 60, 61, 62), meaning that with those 18 questions an answer of *no* would be scored as a one and an answer of *yes* would be scored as a zero. Using a classical scoring method (Heppner et al., 2016), scores for the RIS are determined by adding up the number of positive responses. Higher scores indicate a stronger RI overall.

Procedure

Upon Institutional Review Board approval, the study instruments were uploaded onto the primary investigator's Qualtrics account. At that time, information about the study was uploaded onto the university psychology department's human subject research system (SONA Systems). Once registered on the SONA system, participants were linked to the instruments used for this study through Qualtrics. All participants were asked to read an informational page that briefly described the nature and purpose of the study, and were told that by continuing they were agreeing to participate in the study and could discontinue at any time. Participants consented by selecting "continue" and completed the questionnaire and instruments. After completion, participants were directed to a post-study information page on which they were thanked and provided contact information about the study and the opportunity to schedule a meeting to discuss research findings at the conclusion of the study. No identifying information was gathered from participants. All information was stored on the Qualtrics database.

Results

All analyses were conducted in SAS 9.4 (SAS Institute, 2012). The researchers first used classical methods (e.g., KR_{20} and principal factor analysis) to examine the psychometric properties of the RIS.

Based on the results of the factor analysis, the researchers used results from a one-parameter Rasch analysis to reduce the number of items on the RIS.

Classical Testing

Homogeneity was explored by computing Kuder-Richardson 20 (KR_{20}) alphas. Across all 68 items the internal consistency was strong (.92). Concurrent validity (i.e., construct validity) was examined by looking at correlations between the RIS and the RSES. The overall correlation between the RIS and the RSES was .66 ($p < .001$).

Item Response Analysis

Item response theory brought about a new perspective on scale development (Embretson & Reise, 2000) in that it promoted scale refinement even at the initial stages of testing. Item response theory allows for shorter tests that can actually be more reliable when items are well-composed (Embretson & Reise, 2000). The RIS initially included 68 items. Through Rasch analyses, the scale was reduced to 21 items (items numbered 3, 4, 9, 10, 12, 13, 16, 18, 19, 24, 26, 34, 39, 41, 42, 43, 44, 46, 47, 49, 61).

The final 21 items were selected for their dispersion across location on theta in order to widely capture the constructs. The polychoric correlation matrix for the 21 items was then subjected to a principal components analysis yielding an initial eigenvalue of 11.72. The next eigenvalue was 1.97, which clearly identified the crook of the elbow. Further, Cronbach's alpha for these 21 items was .90. Taken together, these results suggest that the 21-item RIS measures a single factor.

This conclusion was further tested by fitting the items to a two-parameter Rasch model (AIC = 3183.1). Slopes were constrained to unity (1.95), and item location estimates are presented in Table 1. Bayesian *a posteriori* scores also were estimated and strongly correlated with classical scores (i.e., tallies of the number of positive responses [$r = .95$, $p < .0001$]).

Discussion

This scale represents a move from subjective to a more objective assessment of RI. In the future, the scale may be used with other student and non-student populations to better establish its psychometric properties, generalizability, and refinement. Although this study sampled undergraduate students, this scale may be well-suited to use with counseling graduate students and practitioners because items were developed based on a qualitative study with master's-level counseling students and practicing counselors (Jorgensen & Duncan, 2015a).

Additionally, this scale offers another method for assessing student learning and changes that take place for both students and professionals. As indicated by Holden et al. (1999), it is important to assess learning in multiple ways. Traditional methods may have focused on measuring outcomes that reflect a performance-based, rather than a mastery-based, learning orientation. Performance-based learning has been defined as wanting to learn in order to receive external validation such as a grade (Bruning, Schraw, Norby, & Ronning, 2004). Mastery learning has been defined as wanting to learn for personal benefit and with the goal of applying information to reach a more developed personal and professional identity (Bruning et al., 2004).

Based on what is known about mastery learning (Bruning et al., 2004), students with this type of learning orientation experience identity changes that may be best captured through assessing changes in thoughts, attitudes, and beliefs. The RIS was designed to measure constructs that capture internal

changes that may be reflective of a mastery learning orientation. A learner who is performance-oriented may earn an A in a research course but show a lower score on the RIS. The opposite also may be true in that a learner may earn a C in a research course but show higher scores on the RIS. Through the process of combining traditional assessment methods such as grades with the RIS, programs may get a more comprehensive understanding of the effectiveness and impact of their research-related curriculum.

Table 1.

Item location estimates.

RIS Item	Location Estimate
Item 3	-2.41
Item 4	-1.80
Item 10	-3.16
Item 13	-.86
Item 16	-.94
Item 19	-3.08
Item 24	-2.86
Item 9	-1.10
Item 12	.42
Item 18	-2.24
Item 26	-2.20
Item 39	.20
Item 42	-1.28
Item 44	-.76
Item 34	-1.27
Item 41	-.76
Item 43	-1.47
Item 46	-2.03
Item 47	-2.84
Item 49	1.22
Item 61	-.44

Limitations and Areas for Future Research

The sample size and composition were sufficient for the purposes of the initial development and classical testing and item response analysis (Heppner et al., 2016); however, these authors still suggest caution when applying the results of this study to other populations. Endorsements of the participants may not reflect answers of the population in other areas of the country or different academic levels. Future research should sample other student and professional groups. This will help to further establish the psychometric properties and item response analysis conclusions and make the RIS more appropriate for use in other fields. Additionally, future research may examine how scores on the RIS correlate with traditional measures of learning (e.g., grades in individual research courses, collapsed grades in all research courses, research portion on counselor licensure exams).

Conclusion

As counselors-in-training and professional counselors are increasingly being required to demonstrate they are using evidence-based practices and measuring the effectiveness of their services, they may benefit from assessments of their RI (American Counseling Association, 2014; Gladding & Newsome, 2010). CACREP (2016) has responded to increased accountability by enhancing their research and evaluation standards for both master's- and doctoral-level counseling students. The American Counseling Association is further supporting discussions about RI by publishing a recent blog post titled "Research Identity Crisis" (Hennigan Paone, 2017). In the post, Hennigan Paone described a hope for master's-level clinicians to start acknowledging and appreciating that research helps them work with clients in ways that are informed by "science rather than intuition" (para. 5). As the calling becomes stronger for counselors to become more connected to research, it seems imperative that counseling programs assess their effectiveness in bridging the gap between research and practice. The RIS provides counseling programs an option to do exactly that by evaluating the way students are learning and growing in relation to research. Further, the use of this type of outcome measure could provide for good modeling at the program level; in that, the hope would be that it would encourage counselors-in-training to develop both a curiosity and motivation to infuse research practices (e.g., needs assessments, outcome measures, data analysis) into their clinical work.

References

American Counseling Association. (2014). *2014 ACA code of ethics*. Alexandria, VA: Author.
Bruning, R. H., Schraw, G. J., Norby, M. M., & Ronning, R. R. (2004). *Cognitive psychology and instruction* (4th ed.). Upper Saddle River, NY: Pearson Merrill/Prentice Hall.
Council for Accreditation of Counseling & Related Educational Programs. (2016). *2016 CACREP standards.* Retrieved from http://www.cacrep.org/wp-content/uploads/2017/07/2016-Standards-with-Glossary-7.2017.pdf
Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Gladding, S. T., & Newsome, D. W. (2010). *Clinical mental health counseling in community and agency settings* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Hennigan Paone, C. (2017, December 15). Research identity crisis? [Blog post]. Retrieved from https://www.counseling.org/news/aca-blogs/aca-member-blogs/aca-member-blogs/2017/12/15/research-identity-crisis

Heppner, P. P., Wampold, B. E., Owen, J., Thompson, M. N., & Wang, K. T. (2015). *Research design in counseling* (4th ed.). Boston, MA: Cengage Learning.

Hernon, P. & Dugan, R. E. (2004). Four perspectives on assessment and evaluation. In P. Hernon & R. E. Dugan (Eds.), *Outcome assessment in higher education: Views and perspectives* (pp. 219–233). Westport, CT: Libraries Unlimited.

Holden, G., Barker, K., Meenaghan, T., & Rosenberg, G. (1999). Research self-efficacy: A new possibility for educational outcomes assessment. *Journal of Social Work Education*, *35*, 463–476.

Jorgensen, M. F., & Duncan, K. (2015a). A grounded theory of master's-level counselor research identity. *Counselor Education and Supervision*, *54*, 17–31. doi:10.1002/j.1556-6978.2015.00067

Jorgensen, M. F., & Duncan, K. (2015b). A phenomenological investigation of master's-level counselor research identity development stages. *The Professional Counselor*, *5*, 327–340. doi:10.15241/mfj.5.3.327

Kahn, J. H., & Scott, N. A. (1997). Predictors of research productivity and science-related career goals among counseling psychology doctoral students. *The Counseling Psychologist*, *25*, 38–67. doi:10.1177/0011000097251005

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Phillips, J. C., & Russell, R. K. (1994). Research self-efficacy, the research training environment, and research productivity among graduate students in counseling psychology. *The Counseling Psychologist*, *22*, 628–641. doi:10.1177/0011000094224008

Readability Test Tool. (n.d.). Retrieved from https://www.webpagefx.com/tools/read-able/

Reisetter, M., Korcuska, J. S., Yexley, M., Bonds, D., Nikels, H., & McHenry, W. (2004). Counselor educators and qualitative research: Affirming a research identity. *Counselor Education and Supervision*, *44*, 2–16. doi:10.1002/j.1556-6978.2004.tb01856.x

Rowan, N., & Wulff, D. (2007). Using qualitative methods to inform scale development. *The Qualitative Report*, *12*, 450–466.

SAS Institute [Statistical software]. (2012). Retrieved from https://www.sas.com/en_us/home.html

Unrau, Y. A., & Beck, A. R. (2004). Increasing research self-efficacy among students in professional academic programs. *Innovative Higher Education*, *28*(3), 187–204.

Unrau, Y. A., & Grinnell,, R. M., Jr. (2005). The impact of social work research courses on research self-efficacy for social work students. *Social Work Education*, *24*, 639–651. doi:10.1080/02615470500185069

Warden, S., & Benshoff, J. M. (2012). Testing the engagement theory of program quality in CACREP-accredited counselor education programs. *Counselor Education & Supervision*, *51*, 127–140. doi:10.1002/j.1556-6978.2012.00009.x

nbcc