Published in final edited form as:

Educ Res Rev. 2019 June ; 27: 218–228. doi:10.1016/j.edurev.2019.04.002.

Approaches to measuring use of research evidence in K-12 Settings: A systematic review

Jennifer A. Lawlor, Kristen J. Mills, Zachary P. Neal, Jennifer Watling Neal, Camren Wilson, Kathryn McAlindon

Department of Psychology, Michigan State University

Abstract

An increased focus on the use of research evidence (URE) in K-12 education has led to a proliferation of instruments measuring URE in K-12 education settings. However, to date, there has been no review of these measures to inform education researchers' assessment of URE. Here, we systematically review published quantitative measurement instruments in K-12 education. Findings suggest that instruments broadly assess user characteristics, environmental characteristics, and implementation and practices. In reviewing instrument quality, we found that studies infrequently report reliability, validity, and demographics about the instruments they develop or use. Future work evaluating and developing instruments should explore environmental characteristics that affect URE, generate items that match up with URE theory, and follow standards for establishing instrument reliability and validity.

Keywords

Use of research evidence; Measurement; K-12 education

1. Introduction

The use of research evidence¹ (URE) refers to the incorporation of research evidence to make decisions, think about problems and potential solutions, and justify the resolution of problems (Weiss, 1979). Many professional fields have adopted a focus on URE in recent years, attempting to increase the extent to which practitioners incorporate it into their work (Cochrane, 1971, Levant et al., 2006). In order to increase URE, it is necessary to measure it, but instruments for measuring URE are often hard to locate and evaluate. Thus, the generation, testing, and systematic review of instruments for measuring URE has been the focus of recent research efforts (Buchanan, Siegfried, & Jelsma, 2015; Glegg & Holsti, 2010; Estabrooks, Floyd, Scott-Findlay, O'Leary, & Gushta, 2003; Leung, Travena, & Waters, 2014).

In areas like healthcare and psychology, the evidence-based practice movement has worked to shift norms around URE, increasing URE in decision-making among practitioners

¹A variety of other terms are also used to refer to this concept, including: evidence based practice, knowledge utilization, research use, evidence use, evidence informed decision making, and research utilization.

(Cochrane, 1971; Levant et al., 2006). This focus is also present in the field of education where the gap between research and practice is apparent and there have been numerous efforts to promote URE in K-12 education (Behrstock-Sherratt, Drill, & Miller, 2011; Greenberg et al., 2003; Hallfors & Godette, 2002; Authors; Williams & Coles, 2007). The increased focus on URE in K-12 education has led to a proliferation of measures of URE in K-12 education settings. However, to date, there has been no review of these measures to inform education researchers' assessment of URE.

Therefore, this paper has three goals: (1) to summarize the content covered by current quantitative URE instruments used in K-12 education, (2) to evaluate the quality of the measurement instruments, and (3) to establish future directions for measuring URE in this area. To do this, we review the current literature defining URE, establish its importance and current use in K-12 education, and highlight the need for a systematic review of current measurement instruments in this area. We then employ a systematic review approach to assess available instruments of URE in K-12 education. First, we focus on the content themes that the instruments cover. Second, we examine the level of validity and reliability of instruments for measuring URE in K-12 education in order to evaluate the quality of these measures. Finally, we conclude with an assessment of the overall landscape of URE measurement instruments in K-12 education and future directions to improve URE measurement in this setting.

1.1 Use of Research Evidence (URE)

URE has a number of wide-ranging definitions. Some define URE as the presence of a practice culture that includes research evidence (Fernández-Domínguez et al, 2014), while others consider URE to be the use of the 'best' research evidence for a particular situation (Glegg & Holsti, 2010). Still other definitions describe URE as the weighing of contextual factors, stakeholder values, and research evidence in order to make decisions (Shaneyfelt et al, 2006). For our purposes, we define the 'research' component of URE as "systematic data collection and analysis to answer a pre-defined question," typically through the use of the scientific method (Authors). Here, we will use a definition of the 'use' component of URE as outlined by the W.T. Grant Foundation that draws on Weiss (1979):

use of research evidence can happen in many ways and may involve the direct application of research evidence to decision making, conceptual influences on how decision makers think about problems and potential solutions, strategic uses of research to justify existing stances or positions, or imposed uses that require decision makers to engage with research.

(W.T. Grant Foundation, 2016, p. 5)

This definition differs from others in its breadth by including instrumental, conceptual, and symbolic uses of research. Instrumental URE refers to use of information that can be directly documented and is often very specific or task-oriented (Caplan, Morisson, & Slambaugh 1975; Rich, 1975). For example, instrumental URE occurs when a principal adopts a new math curriculum after reading a study showing that it had positive student outcomes. Conceptual URE is characterized as being broader and influencing the way stakeholders think about an issue (Caplan, Morrison, & Slambaugh, 1975). For example,

conceptual use could include school administrators' use of a literature review to guide their thinking about an area of practice on which they are working to improve. Such a document may provide a frame to begin crafting district-level policies that align with research. Finally, symbolic URE refers to situations in which research evidence is used to support or justify those decisions that have already been made (Knorr, 1977). For example, this could occur in situations where teachers or administrators seek out research evidence to provide support for a practice they are currently implementing. Although frequently linked to Weiss (1979), this three-part classification of types of research use has been widely adopted in the theoretical and empirical literature (e.g. Pelz, 1978; Beyer & Trice, 1982; Estabrooks, 1999) and foundations supporting work in this area (e.g. W. T. Grant Foundation, 2016).

1.2 URE in K-12 Education

Because school systems throughout the world have been moving toward incorporating URE into their work, measuring URE is particularly important in K-12 education. In the United States, for example, policies like the No Child Left Behind act (NCLB, 2011) and the Every Student Succeeds act (ESSA, 2015) have mandated URE in public school settings, tying school compliance and progress with funding. This means that the pressure to implement research evidence can also have implications for education funding in U.S. education contexts. Both NCLB and ESSA provide federal expectations for research use in K-12 settings, but involve different conceptions of research. NCLB is focused on the concept of scientifically rigorous research, with a particular focus on random-assignment experiments, while ESSA focuses on the concept of evidence-based programming and includes multiple tiers of evidence based on the rigor of the research design implemented in the study (NCLB, 2001; ESSA, 2015; Farley-Ripple, May, Karpyn, Tilley, & McDonough, 2018).

The push for URE has also been seen in the United Kingdom with the promotion of a 'self-improving school system,' which has led to many schools adopting action research initiatives as part of their improvement process. These initiatives bring research directly into schools, making it possible for educators to engage in research most useful to their context (Roberts, 2015). Similar efforts to increase URE have also occurred in Australia (Blackmore, 2002) and Canada (Campbell, Pollock, Briscoe, Carr-Harris & Tuters, 2017).

At present, there are a number of measures of K-12 URE. However some measures are difficult to implement due to constraints in K-12 settings and they are often challenging to find. Many current approaches to studying URE in education rely on qualitative designs that can be time and resource intensive (e.g., Miretzky, 2007; Nicholson-Goodman & Garman, 2007). Similar to professionals in many fields, individuals working in K-12 settings often have full schedules that make it difficult to establish time for things like interviews or focus groups. Therefore, quantitative measures that can be completed quickly and with limited resource strain are often more useful. However, scholars have noted challenges in finding appropriate instruments for measuring K-12 URE and instead are creating their own instruments. For example, Jimerson (2016) indicates "the educational landscape is fraught with poorly-conceived survey instruments" (p. 5). Thus, bringing together the piecemeal literature regarding quantitative measurement in this area can help scholars identify appropriate instruments for their needs.

There are a number of systematic reviews of URE measurement instruments in other disciplines like nursing and occupational therapy (Buchanan, Siegfried, & Jelsma, 2015; Estabrooks, Floyd, Scott-Findlay, O'Leary, & Gushta, 2003; Glegg & Holsti, 2010; Leung, Travena, & Waters, 2014). These reviews provide information regarding the dimensions of URE for which there are instruments within their fields as well as the psychometric properties of the available instruments. Knowing this information can be useful for helping researchers to identify a psychometrically sound instrument that evaluates the specific dimension of URE relevant to their research questions. They also often provide information about future directions for addressing gaps in measurement in order to support continuing work to improve the instruments in their area. While these types of reviews are common in medical fields, there are currently no systematic reviews of instruments for measuring URE in the field of K-12 education, which represents a unique context for the measurement of URE and where unique measurement challenges may arise. Generating such a review will be useful for addressing the challenges associated with finding and evaluating appropriate quantitative instruments. In addition, it will support future work that addresses gaps in the dimensions current URE instruments measure.

2. Methods

We conducted a systematic review of available instruments that assess K-12 URE. Specifically, we wanted to find articles containing instruments measuring knowledge, skills, attitudes, implementation, barriers, facilitators of URE for this review. We then extracted the instruments from the articles. Below, we discuss the process we used to search for and select articles (see figure 1), as well as the process of extracting instruments from these articles and evaluating the instruments' content and quality.

2.1 Search & Inclusion Process

Figure 1 illustrates the process we used to systematically identify instruments designed to quantitatively measure URE in K-12 settings. This process relied primarily on the ProQuest database because it covers a breadth of scholarly publications (almost 8,000), including all publications in the Education Resources Information Center (ERIC). The size of ProQuest combined with the variety of publications regarding education ensured that the articles returned would reflect both education-specific journals as well as those from fields adjacent to education. While Google Scholar is comprehensive in its indexing, its search function is not as sophisticated as ProQuest for this type of search. For example, in ProQuest we were able to apply 'wild card' words within our searches, using the term "measur*" to search for anything that included "measure," "measurement," or any other variant on the root "measur." Google Scholar does not yet have the capability to complete this type of search. However, during the search process we became aware that a journal from the American Education Research Association, *AERA Open* is only partially indexed in ProQuest. Thus, we conducted a modified Google Scholar search to ensure that we were able to capture articles from this education-focused journal. To capture this, we conducted a specific search for terms related to URE and measurement within *AERA Open*.

In the first step, we used ProQuest to locate all articles that include (1) one of the following keywords related to URE in their title: “evidence based practice,” “knowledge utilization,” “use of research,” “using research,” “research use,” “research evidence,” “evidence use,” “evidence informed decision making,” “data use,” “data based decision making,” “research utilization”, (2) one of the following words related to measurement in their abstract: *scal*², *survey**, *valid**, *measur**, *instrument**, and (3) one of the following field keywords in their abstract: “social work*”, “mental health,” *school**, *educat**, “public health”. We included these diverse keywords in order to get a broad spectrum of studies that might be relevant to K-12 education, so that we would have a rich pool to refine in the next step. Professionals working in K-12 education come from a variety of backgrounds and using diverse keywords like mental health and public health increased the number of results from fields like school nursing, school psychology, and occupational therapy. In addition, limiting the field-specific keywords helped focus results on the interests for this study, rather than returning articles primarily about medical settings, where the majority of URE measurement research takes place. In this step, the pool included 551 articles.

In step two, two members of the research team read the title and abstract for each of the articles in the pool and came to consensus about tentative article inclusion. Based on their title and abstract, articles were included if they appeared to contain a: (a) an instrument measuring URE, (b) that is quantitative, (c) in a K-12 education setting. In instances where it was not clear whether the article had met inclusion criteria, it was included and reassessed when team members read the full articles. At this step, the pool was reduced to 46 articles. Articles were removed at this step for a number of reasons. We excluded 261 articles from the pool because they did not contain an instrument that measures URE. Another 11 were excluded because they did not include a quantitative instrument. Finally, 233 articles were excluded because they did not measure URE in a K-12 educational setting. Some articles were excluded for multiple reasons (e.g. they lacked a quantitative measure of URE in a K-12 setting).

In step three, two team members read each of the included papers in full to evaluate whether the full text met the inclusion criteria. At this step, we filtered out twelve articles, reducing the size of our pool to 34. Seven articles were excluded because they did not contain an instrument that measured URE and five were excluded because they did not include a quantitative instrument. At this stage, we also were able to review the content of the scales and identified data use scales as measuring a separate construct from URE. This matches up with literature suggesting that data use can be understood as its own subfield of educational data sciences (Piety, Hickey, & Bishop, 2014). Thus, we removed 11 articles that focused specifically on data use, reducing the size of the pool to 23.

In step four, the team members then transformed the pool of articles into a pool of instruments (step 4 in Figure 1). Because some articles in the pool reported on multiple

²The asterisk symbol here represents terms that were used as wildcards during the search process, meaning we looked for responses that used those terms either on their own or as part of another word (i.e., scale, scaled, and scales would all come up when using the search term *scale**).

instruments, while other articles reported on the same instruments, we extracted 18 unique instruments from the pool of 23 articles.

Finally, in step 5, we examined items referenced by these 23 articles and performed a Google search for each named instrument, to identify any additional materials that provide measurement information about applications of these instruments in K-12 settings.³ We included this final step to ensure that our review of instruments' measurement properties included any details contained in materials not already captured by our search. This led to the inclusion of 1 additional article, 1 unpublished dissertation, and 1 technical report, thus in total we report on 18 instruments drawing on information from 26 sources.

2.2 Data extraction process

Two coders extracted the following information about the instruments from each source: (1) the dimensions or features of URE the authors described the instrument as intended to measure, (2) evidence of the instrument's validity and reliability, (3) whether the measures came from or were adapted from another source, and (4) study characteristics including response rate, participant demographics, and the location(s) of the study. To establish instrument quality, we consider validity to be evidence that an instrument effectively measures the construct (here, URE) that it is intended to assess, and reliability to be evidence that an instrument is stable over time and/or across items. There are multiple types of validity, including construct, external, generalizability, content, substantive, and structural. We extracted any instance in which a study explicitly discussed the validity of the instrument along with any evidence presented to support the validity claim. Likewise, we extracted data on several types of instrument reliability – test-retest reliability, scale reliability, and subscale reliability – if the reliability was computed using the sample in the article, but not if reliability was reported from a different sample. We chose this approach because this pool focuses specifically on K-12 settings and we wanted to include reliabilities that matched the settings in which these studies were conducted. The two coders compared the extracted data for inconsistencies and came to consensus about any differences.

2.3 Content coding & theming process

To identify the content themes covered by the instruments, the two coders undertook a process of coding and theming. As described in section 2.2, the dimensions or features of URE the authors described the instrument as intended to measure were first identified and extracted. For example, Brown & Zhang (2017) report that their instrument was intended to measure, among other things, “teacher capacity to engage in and with research” (p. 281). Second, two coders independently assigned codes that described the content of the extracted dimensions. For example, the quote above was coded as measuring the dimension of “capacity.” The coders solicited feedback from the other authors on the paper to refine and re-apply codes, and iterated this process until they reached consensus. Third, the coders

³The requirement that items included in step 5 only involve applications in K-12 settings is important. For example, by step 4 our review included the Adapted Fresno Test and the BARRIERS scale, which each have large measurement literatures. However, we do not include these vast literatures in step 5 because they have been exclusively focused on applications of these instruments with non-school-based occupational therapists and nurses, respectively. Additionally, reviews of these instruments in these contexts already exist (Glegg & Holsti, 2010; Kajermo et al., 2010)

aggregated codes into larger content themes. For example, the *capacity* code was aggregated, together with codes relating to *confidence* in using research evidence and codes relating to *perceptions* of URE, into a theme of “user characteristics.” Again, the coders solicited feedback from the other authors and refined the aggregation of codes into themes until they reached consensus.

3. Results

After extracting data on each of the instruments included in the review, we assessed the characteristics of the studies and the URE content themes (see table 1), and the quantitative indicators of instrument quality (see table 2). The studies included in our review applied URE instruments in a diverse group of educators, including administrators, teachers, and school staff members. Most frequently, the instruments were used with principals ($n=5$, 27.77%), and teachers ($n = 4$, 22.22%). Participants also occupied a number of other education-related roles, including school psychologists, school nurses, and librarians. The studies implementing these instruments were conducted in the United States ($n = 11$) and Canada ($n = 2$), with single studies from England, Germany, Israel, Scotland, and Wales.

3.1 Content themes in URE measurement

Our first research question considers the content themes covered in existing URE instruments. We found three overarching themes in the content covered by the instruments: user characteristics (UC, $n = 16$, 88.88%), environmental characteristics (EC, $n = 11$, 61.11%), and implementation and practices (I&P, $n = 14$, 77.77%; see table 1). Most instruments ($n=14$, 77.77%) included content representing multiple themes.

The first major content theme focused on the characteristics of the users of research evidence, which included perceptions and attitudes related to research evidence, capacity for using research, and confidence locating, evaluating, or using research. Perceptions and attitudes toward URE was the most frequently coded dimension within this theme ($n = 11$, 61.11%), and included the presence of items such as “Current research is useful in the day-to-day management of my clients” (Cahill et al., 2013). Capacity for use was also a frequently observed dimension within this theme ($n = 9$, 50%). The Adapted Fresno Test, described in Cahill, Egan, Wallingford, Huberlee, and Dess-McGuire (2013) is an example of an instrument measuring capacity. It provides occupational therapy scenarios and asks participants to do things like interpret validity of a study and evaluate statistical significance of the findings. Finally, some instruments aimed to measure users’ confidence in locating, evaluating, or using research ($n = 3$, 16.66%). For example, the instrument presented in Williams and Coles (2007a, b) includes questions about how confident participants were in their ability to complete several URE-related tasks: identifying information needs, finding research information, and evaluating the information they find.

The second major content theme focuses on environmental characteristics, including information sources, effective structures for URE, and school culture and research promotion. Instruments assessing information sources evaluate the extent to which participants have access to sources that disseminate information about research and the sources they use to access information ($n = 6$, 33.33%). For example, Massell, Goertz, and

Barnes (2017) include questions that evaluate where individuals find the research relevant to their work and whether those sources are internal or external to their organization. Instruments measuring the extent to which organizations provided an effective structure for URE included factors such as offering the time, resources, events, or capital to make URE possible ($n = 5$, 27.77%). For example, the instrument in Lysenko, Abrami, Bernard, & Dagenais (2015) asks participants to rate the extent to which “available facilities and technology” influences their URE. Finally, instruments measuring school culture and research promotion evaluate the extent to which school culture is attuned to URE, promote URE, make URE a norm, and facilitate evidence use ($n = 4$, 22.22%). For example, the instrument described by Brown and Zhang (2017) asks respondents whether “*My school encourages me to use research findings in order to improve my practice*” (Brown & Zhang, 2017).

The third major content theme focuses on implementation and practices, including current implementation and factors affecting implementation. The assessment of current implementation aims to measure the actual current use of research evidence, evidence-based practices, or evidence-informed practice at the individual level ($n = 10$, 55.55%). For example, the School-Nurse Evidence Based Practice Questionnaire aims to capture current implementation by asking questions about how frequently school nurses use a series of evidence-based interventions in their practice (Adams, 2009; Adams & Barron, 2009; Adams & Barron, 2010). In contrast, the assessment of factors affecting implementation aims to measure the facilitators and barriers to URE ($n = 6$, 33.33%). For example, Hemsley-Brown and Oplatka (2005) examine barriers to research use by asking respondents whether they endorse statements such as “*Research reports are not published fast enough*”.

3.2 Instrument quality assessment

To evaluate instrument quality, we focused on evidence of ease-of-use, validity and reliability extracted from each article (see table 2). Instruments’ ease of use can be evaluated in part from the sizes of the samples within which they have been collected, and the response rates achieved in those populations as reported in each article. Participant sample sizes for studies using these instruments varied from 2 to 2425; however, many reported findings from samples of less than 100 participants ($n = 6$, 33.33%). Similarly, participant response rates varied from 11.4% to 100%, but 27.8% of the instruments ($n = 5$) have been used only in studies where the response rate is unknown or unreported.

Sources on 7 of the 18 instruments (38.8%) report some kind of validity, however in many cases little evidence was provided to substantiate claims of validity. Studies have claimed to demonstrate the face or content validity of six instruments, typically by expert review of the instrument or by a pilot administration of the instrument. Studies also claimed to demonstrate the structural validity of three instruments, which exists when sets of items tap specific theoretically-defined aspects of a construct, by using either principal components analysis or exploratory factor analysis. Finally, studies claimed to demonstrate three instruments’ convergent or discriminant validity by examining the instrument’s association with other variables, including education, attitudes, and role. In the discussion section we

further consider the types of validity expressed in the sources against popular standards for validity.

Sources on 5 of the 18 instruments (27.7%) reported scale or subscale reliability in the form of inter-item correlations using Cronbach's α ; this was the only types of reliability reported in the sources included in the review. Overall scale reliability was reported for only two instruments (0.82 and 0.94), while subscale reliability was reported for four instruments and ranged from 0.531 to 0.93.

Information about both reliability and validity was presented for only three instruments (16.6%), which coincidentally all had specific names: the School Nurse Evidence Based Practice scale (SN-EBP; Adams, 2007, 2009; Adams & Barron, 2009, 2010), the Questionnaire about the Use of Research-Based Information (QURBI; Lysenko et al., 2014, 2015), and the Survey of Practitioners' Use of Research (SPUR; Penuel et al., 2016, 2017).

4. Discussion

The purpose of this study was to assess the content and quality of current URE measurement instruments for K-12 education settings, and establish future directions for URE measurement in this area. Our findings illuminate the current nature of measurement options as well as critical areas for improvement in both developing instruments and reporting on them.

4.1 Instrument Content

Through a process of content analysis, we identified three content themes in the dimensions of URE covered by the instruments in our review: user characteristics, environmental characteristics, and implementation and practices. These themes map onto Tseng's (2012) conceptual framework of URE, which highlights the importance of understanding both the users of URE as well as the environments that affect URE. Indeed, the themes provide good high-level coverage of these aspects of URE, with instruments measuring characteristics of people (i.e. users), places (i.e. environments), and the interactions among them (i.e. implementation and practice). The user characteristics theme covers the features of people who engage in URE, including dimensions like their capacity for URE and their attitudes or perceptions toward URE. The environmental characteristics theme covers features of the context in which URE happens, including dimensions like school culture and information sources for URE. The implementation and practice theme covers the intersection between users and their environment, focusing on features like how frequently users currently implement research or evidence-based practices in their settings and barriers to an individual's implementation of these practices in their setting. Although these major themes provide good coverage of URE concepts, there are still areas for further exploration in instrumentation. Specifically, there is more room to consider environmental and contextual factors in URE measurement and to create a closer link between theory and measurement.

Among the major themes, user characteristics were most likely to be represented in the instruments that we reviewed. This highlights the importance of individual users for URE

in K-12 settings, but may also suggest an overreliance on user characteristics at the expense of examining the role of environmental characteristics when studying URE. Although individual characteristics may be important for some aspects of URE, the literature suggests the importance of environmental context in promoting behaviors like URE (Domitrovich et al., 2008; Ringeisen, Henderson, & Hoagwood, 2003). For example, Ringeisen, Henderson, and Hoagwood (2003) point to organizational characteristics like resource availability and organizational climate as significant features influencing URE related to mental health in schools. Future work assessing these environmental characteristics and aligning measurement instruments to capture them can be important for understanding and establishing research-friendly settings for educators.

Although these instruments of the major themes surrounding URE, only two – QURBI (Lysenko et al., 2014, 2015) and SPUR (Penuel et al., 2016, 2017) – also explicitly map onto Weiss's (1979) theory about the types of use. This widely used theory outlined three types of URE – conceptual, instrumental, and symbolic – which are potentially relevant to each of the themes discussed in section 3.1. Conceptual URE might focus on educators' use research to increase general knowledge about priority areas in their work, instrumental URE on educators' use of research to inform a specific decision, and symbolic URE on educators' use of research to justify decisions already made.

Educational researchers have long recognized the context specificity of measures as important for interpreting their validity and generalizability (e.g., Henson, 2002; Messick, 1984; Moss, 1992). K-12 educational settings are marked by significant variation in their stakeholders, geographic locations, and the situations where stakeholders may encounter that involve the use of research evidence. Thus, it is important to view our findings in light of this variation, recognizing that some instruments are more general while others may be context specific. For example, while a plurality of instruments included in this review were used to study principals or teachers (N = 8), some instruments focused their items on more unique stakeholders within K-12 settings. For example, the SN-EBP is specifically focused on the context of nursing within K-12 settings, although it measures user characteristics, environmental characteristics, and implementation and practice (Adams, 2009; Adams & Barron, 2009; Adams & Barron, 2010). While it measures URE across multiple themes, it would not be appropriate to use the SN-EBP every educational stakeholder group. In contrast, the QURBI and SPUR were focused on more general populations of educators, making them potentially more widely applicable. However, even here there are subtle differences: the QURBI was developed and has been tested in samples dominated by front-line practitioners such as teachers, while the SPUR was developed and tested in samples dominated by top-level executives such as district superintendents. With respect to geographic location, a plurality of instruments included in this review were used in the United States or Canada (N = 13), and therefore may be most useful in educational settings in North America. Finally, with respect to situations, some instruments ask educators to report on their URE generally (e.g., Lysenko et al., 2015) while others ask educators to report on their URE in a particular situation (e.g., Cahill et al, 2013). These latter situation-specific measures of URE are more fine-grained, but may lack generalizability. Thus, as researchers consider selecting a URE measurement instrument, it is important to consider

whether the instrument was developed for the specific population, location, and level of specificity under investigation.

We suggest a couple of future directions for research on the content of URE instruments. First, researchers should further develop instruments measuring the environmental context of URE; these will be necessary to establish the extent to which context influences K-12 URE. Second, researchers should generate a closer link between theory and measurement for URE by applying Weiss' (1979) types of URE as a guide for developing new instruments or adapting current instruments. Third, researchers should investigate whether existing instruments can be used (with or without adaptation) outside North America, and if not, should develop new instruments suitable for other geographic contexts.

4.2 Instrument Quality

Turning to measurement quality, while the instruments in our review measure a variety of dimensions of URE in K-12 education, few occur in studies with large samples or high response rates. Low or missing response rates may indicate challenges with the feasibility of implementing an instrument. In addition, the sources provided little demographic information and the information provided was inconsistent across sources. Without information about demographics at the individual and setting level, we cannot assess instrument generalizability. Specifically, researchers can report on things like race, gender, position, position tenure, and school context (e.g., rural or urban, economic factors).

Similarly, few studies using these instruments offer evidence of validity (38.8%) or reliability (27.77%). Because many studies did not report evidence of the validity or reliability, it is difficult to determine the quality of many of the URE measures we reviewed. Even when evidence of validity or reliability was presented, it was limited in nature. For example, only evidence of face or content validity were provided for some instruments reporting on validity (e.g., review of items from the target population or topical experts for content validity). This evidence of validity is weak in consideration of existing guidelines concerning evidence for validity. For example, Messick (1995) indicates that construct validity has six aspects, including: content, substantive, structural, generalizability, external, and consequential, each of which has unique evidence requirements. External validity, for example, requires showing that the instrument has evidence of convergence, or, correlation with related constructs as well as divergence, or the lack of a correlative relationship with concepts unrelated to the construct of interest.

Similarly, evidence of reliability was often insufficient to evaluate quality. For example, two sources provides whole scale reliability for instruments without providing their subscale reliabilities, despite the instruments being designed to measure multiple dimensions of URE. In addition, two instruments have evidence of reliability but not of validity. Without evidence of validity, it is impossible to assess whether a reliable instrument is measuring the construct of interest.

4.3 Recommended instruments

Three the instruments included in the review stood out for their content coverage and measurement quality, and are recommended for future work: SN-EBP, QURBI, and SPUR.

This recommendation is informed by several features of these instruments. First, they each cover all three URE themes of users, environments, and implementation and practices, thereby providing adequate content coverage. In addition, both the QURBI and SPUR are also explicitly linked to, and are designed to separately assess, each of Weiss's (1979) theoretically distinct types of research use. Second, they have been employed in large samples ranging from a few hundred for the SN-EBP to several thousand for the QURBI, in each case with response rates greater than 50%, suggesting they are feasible to collect in K-12 school settings. Third, they provide evidence of both reliability and validity when collected in K-12 school settings, and the evidence for validity includes content, structural, and external (i.e. convergent or discriminant) validity. Finally, each complete instrument is readily available in either a published article (Adams & Barron, 2010; Lysenko et al., 2015) or via a project website (<http://www.ncrpp.org>), and can be used by other researchers at no cost.

It is important to note that all three of the instruments were developed and tested in a North American context, and may not be suitable for measuring URE in other places. In contrast, they do offer variation in the populations for which they are suitable. The SN-EBP is a specialized instrument designed for use with school nurses, and with minor adaptation may also be suitable for other school-based health professionals. The QURBI and SPUR were designed for use with more general populations of educators, with the QURBI more focused on front-line practitioners such as teachers, and the SPUR more focused on executive roles such as superintendents and principals.

4.4 Limitations and future directions

Our findings should be interpreted in light of some limitations. First, the keywords we used in our search process may have missed some articles that included URE instruments. However, we have tried to ensure strong coverage of URE instruments by using multiple common terms for URE, using a variety of field-related terms to capture articles in fields that may have published instruments related to K-12 URE, and searching a large database that covers a wide range of journals relevant to this topic. Second, the scope of our review focused only on quantitative instruments in published literature on K-12 education. Thus, there may be additional instruments that are qualitative, written about in unpublished sources like technical reports, or instruments being implemented outside of K-12 education that would be appropriate for measuring URE in this context. Future studies can build upon our review by expanding their scope to include instruments in qualitative, unpublished, and non-K-12 literature. Future research may also explore additional search engines that index relevant journals not included in ProQuest.

Although our inclusion criteria focused on quantitative instruments, we did not limit the instruments to self-report surveys. However, the instruments included in the articles we reviewed exclusively used self-report surveys as a way of collecting information about URE. Self-report surveys can be limiting because they may be reactive or rely heavily on participant recall. Having a variety of approaches to measuring URE can help provide more options for researchers to collect high quality data, and highlights another future direction for research on URE measurement. For example, a fully structured observation of

research use in meetings could demonstrate how often various uses of research are invoked, while archival approaches to measuring URE may also help to understand the frequency and kinds of research used to make decisions in schools without being reactive (Neal et al., 2019).

Our findings about the measurement quality of existing URE instruments also point to several ways researchers can improve measurement in this area. First, researchers should consider the wide variety of instruments available before choosing to create their own instrument. They can look to systematic reviews to quickly establish available options. While this review provides an overview of instruments specific to K-12 settings, there are systematic reviews that include instruments for measuring URE across a number of fields (Buchanan, Siegfried, & Jelsma, 2015; Estabrooks, Floyd, Scott-Findlay, O'Leary, & Gushta, 2003; Glegg & Holsti, 2010; Leung, Travena, & Waters, 2014). These may point to instruments that can be adapted for use in a K-12 setting. Second, when researchers use URE measurement instruments, they should report reliability and validity using established standards in the field. The American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (2014) established a set of joint standards for educational and psychological testing. Implementing these standards when testing instruments and publishing the findings ensures that researchers are using consistent standards and will allow for comparisons among competing instruments. Third, researchers should report on response rate and participant characteristics. Specifically reporting these characteristics will make it possible for others assess feasibility and generalizability of a given instrument. Fourth, researchers should publish or make available the exact wording in the instruments they use or develop. With the exact wording, it is easy to assess the appropriateness of an instrument for a particular research question and to consider instrument feasibility.

To date, there have been limited attempts to systematically evaluate the available instruments for measuring URE. This systematic review fills this gap by assessing the content and quality of existing instruments designed to measure URE in K-12 settings, and by establishing future directions for research. We found that the available instruments measure multiple content themes within URE, but that evidence supporting the reliability and validity of these instruments is limited. Nonetheless, high-quality specialized (SN-EBP in Adams & Barron, 2010) and general (QURBI in Lysenko et al., 2015; SPUR in Penuel et al., 2017) exist. These instruments provide a promising starting point for meaningfully measuring URE in K-12 settings and future research will help solidify measurement instruments to move forward scholarship in this area.

References

- Adams S (2007). Understanding the variables that influence translation of evidence-based practice into school nursing. Unpublished dissertation, University of Iowa.
- Adams S (2009). Use of evidence-based practice in school nursing: Survey of school nurses at a national conference. *The Journal of School Nursing*, 25, 302–313. [PubMed: 19363108]
- Adams S & Barron S (2010). Development and testing of an evidence-based questionnaire for school nurses. *Journal of Nursing Measurement*, 18, 3–25. [PubMed: 20486474]

- Adams S & Barron S (2009). Use of evidence-based practice in school nursing: Prevalence associated variables, and perceived needs. *Worldviews on Evidence-Based Nursing*, 6, 16–26. [PubMed: 19302544]
- AERA, APA & NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Behrstock-Sherratt E, Drill K, Miller S (2011). *Is the supply in demand? Exploring how, when and why teachers use research* (revised ed). Washington, DC: American Institutes for Research.
- Beyer JM, & Trice HM (1982). The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly*, 27, 591–622.
- Blackmore J (2002). Is it only ‘What works’ that ‘Counts’ in New Knowledge Economies? Evidence-based Practice, Educational Research and Teacher Education in Australia. *Social Policy and Society*, 1, 257–266.
- Brown C, & Zhang D (2017). Accounting for discrepancies in teachers’ attitudes towards evidence use and actual instances of evidence use in schools. *Cambridge Journal of Education*, 47, 277–295.
- Brown C, & Zhang D (2016a). Is engaging in evidence-informed practice in education rational? What accounts for discrepancies in teachers’ attitudes towards evidence use and actual instances of evidence use in schools? *British Educational Research Journal*, 42, 780–801.
- Brown C & Zhang D (2016b). Un-rational behaviour? What causes discrepancies between teachers’ attitudes towards evidence use and actual instances of evidence use in schools? *Journal of Educational Administration*, 54, 469–491.
- Buchanan H, Siegfried N, & Jelsma J (2016). Survey Instruments for Knowledge, Skills, Attitudes and Behaviour Related to Evidence-based Practice in Occupational Therapy: A Systematic Review. *Occupational Therapy International*, 23, 59–90. [PubMed: 26148335]
- Cahill SM, Egan BE, Wallingford M, Huber-lee C, & Dess-mcguire M (2013). Results of a School-Based Evidence-Based Practice Initiative. *The American Journal of Occupational Therapy*, 69, 1–6.
- Campbell C, Pollock K, Briscoe P, Carr-Harris S & Tuters S, 2017. Developing a knowledge network for applied education research to mobilise evidence in and for educational practice. *Educational Research*, 59, 209–227.
- Caplan N, Morrison A & Slambaugh RJ (1975). *The use of social science knowledge in policy decisions at the national level*. Ann Arbor, MI: Institute for Social Research.
- Cochrane AL (1971). *Effectiveness and efficiency: Random reflections on health services*. Cambridge: Royal Society of Medicine Press
- Cooper A, & Levin B (2010). Using research in secondary schools: Education leaders respond. *Education Canada*, 50, 58–62.
- Cooper A, Levin B (2013). Research use by leaders in Canadian school districts. *International Journal of Education Policy & Leadership*, 8, 1–15.
- Demski D, & Racherbäumer K (2015). Principals’ evidence-based practice – findings from German schools. *International Journal of Educational Management*, 29, 735–748.
- Domitrovich CE, Bradshaw CP, Poduska JM, Hoagwood K, Buckley JA, Olin S, ... Ialongo NS (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, 1, 6–28. [PubMed: 27182282]
- Estabrooks CA (1999). The conceptual structure of research utilization. *Research in Nursing and Health*, 22, 203–216.
- Estabrooks CA, Floyd JA, Scott-Findlay S, O’Leary KA, & Gushta M (2003). Individual determinants of research utilization: A systematic review. *Journal of Advanced Nursing*, 43, 506–520. [PubMed: 12919269]
- Every Student Succeeds Act (ESSA) of 2015, 20 U.S.C.A. § 6301 *et seq.* (U.S. Government Publishing Office, 2015).
- Farley-Ripple E, May H, Karpyn A, Tilley K, & McDonough K (2018). Rethinking Connections Between Research and Practice in Education: A Conceptual Framework. *Educational Researcher*, 47, 1–11.

- Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Oliva-Pascual-Vaca A, Salinas-Bueno I, & Pedro-Gómez JE (2014). Validity and reliability of instruments aimed at measuring Evidence-Based Practice in Physical Therapy: a systematic review of the literature. *Journal of evaluation in clinical practice*, 20, 767–778. [PubMed: 24854712]
- Finnigan KS, Daly AJ, & Che J (2013). Systemwide reform in districts under pressure: The role of social networks in defining, acquiring, using, and diffusing research evidence. *Journal of Educational Administration*, 51, 476–497.
- Glegg SMN, & Holsti L (2010). Measures of knowledge and skills for evidence-based practice: A systematic review. *Canadian Journal of Occupational Therapy*, 77, 219–232.
- Greenberg MT, Weissberg RP, O'Brien MU, Zins JE, Fredericks L, Resnik H, & Elias MJ (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, 58, 466–474.
- Hallfors D, & Godette D (2002). Will the 'Principles of Effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Education Research*, 17, 461–470. [PubMed: 12197591]
- Hemsley-Brown J & Oplatka I (2005). Bridging the research-practice gap: Barriers and facilitators to research use among school principals from England and Israel. *International Journal of Public Sector Management*, 18, 424–446.
- Henson RK (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, 37, 137–150.
- Jimerson JB (2016). How are we approaching data-informed practice? Development of the survey of data use and professional learning. *Educational Assessment, Evaluation and Accountability*, 28, 61–68.
- Kajermo KN, Boström A-M, Thompson DS, Hutchison AM, Estabrooks CA, & Wallin L (2010). The BARRIERS scale – the barriers to research utilization scale: A systematic review. *Implementation Science*, 5, 32. [PubMed: 20420696]
- Knorr KD (1977). Policymakers' use of social science knowledge: Symbolic or instrumental. In Wiss CH (Ed.) *Using Social Research in Public Policy Making* (165–182). Lexington: Lexington Books.
- Leung K, Trevena L, & Waters D (2014). Systematic review of instruments for measuring nurses' knowledge, skills and attitudes for evidence-based practice. *Journal of Advanced Nursing*, 70, 2181–2195. [PubMed: 24866084]
- Levant RF, Barlow DH, David- KW, Hagglund KJ, Hollon SD, Johnson JD, ... Directorate P (2006). Evidence-Based Practice in Psychology. *American Psychologist*, 61, 271–285.
- Levin B, Cooper A, Arjomand S, & Thompson K (2011). Can simple interventions increase research use in secondary schools? *Canadian Journal of Educational Administration and Policy*, 126, 1–29.
- Lysenko LV, Abrami PC, Bernard RM, Dagenais C, & Janosz M (2014). Education research in educational practice: Predictors of use. *Canadian Journal of Education*, 37, 2–26.
- Lysenko LV, Abrami PC, Bernard RM, & Dagenais C (2015). Research use in education: An online survey of school practitioners. *Brock education Journal*, 25, 35–54.
- Massell D, Goertz ME, & Barnes CA (2017). State education agencies' acquisition and use of research knowledge for school improvement. *Peabody Journal of Education*, 87, 609–625.
- Messick S (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Messick S (1984). *The psychology of educational measurement*. Princeton, NJ: Educational Testing Service.
- McKee WT, Witt JC, Elliott SN, Pardue M, & Judycki A (1987). Practice informing research: A survey of research dissemination and knowledge utilization. *School Psychology Review*, 16, 338–347.
- Meline T, Paradiso T (2003). Evidence-based practice in schools: Evaluating research and reducing barriers. *Language, Speech & Hearing Services in Schools*, 34, 273–283.
- Miretzky D (2007). A view of research from practice: Voices of teachers. *Theory into Practice*, 46, 272–280.

- Moss PA (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229–258.
- Neal ZP, Lawlor JA, Neal JW, Mills K, & McAlindon K (2019). Just google it: Measuring schools' use of research evidence with internet search results. *Evidence and Policy*, 15, 103–123. [PubMed: 31073322]
- Nicholson-Goodman J & Garman NB (2007). Mapping practitioner perceptions of 'It's research based': Scientific discourse, speech acts and the use and abuse of research. *International Journal of Leadership in Education*, 10, 283–299.
- No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.* (West 2003).
- Pelz DC (1978). Some expanded perspectives on use of social science in public policy, Pp. 436–357 in Yinger M, Cutler SJ (eds), *Major social issues: A multidisciplinary view*, New York: Free Press.
- Penuel WR, Briggs DC, Davidson KL, Herlihy C, Sherer D, Hill HC, Farrell C, Allen A (2016). Findings from a national study on research use among school and district leaders. Technical Report No. 1, National Center for Research in Policy and Practice.
- Penuel WR, Briggs DC, Davidson KL, Herlihy C, Sherer D, Hill HC, Farrell C, Allen A (2017). How school and district leaders access, perceive, and use research. *AERA Open*, 3, 1–17.
- Piety PJ, Hickey DT, & Bishop M (2014). Educational Data Sciences: Framing Emergent Practices for Analytics of Learning, Organizations, and Systems. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 193–202.
- Rich RF (1975). Selective utilization of social science related information by federal policy-makers. *Inquiry*, 12, 239–245. [PubMed: 126957]
- Ringeisen H, Henderson K, & Hoagwood K (2003). Context matters: Schools and the “research to practice gap” in children’s mental health. *School Psychology Review*, 32, 153–168.
- Roberts C (2015). Impractical research: Overcoming the obstacles to become an evidence-informed school. In Brown C (ed.), *Leading the use of research and evidence in public schools*. London: Institute of Education Press.
- Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, & Whelan C (2006). Instruments for Evaluating Education in Evidence-Based Practice. *Journal of the American Medical Association*, 296, 1116–1127. [PubMed: 16954491]
- Sortedahl C (2012). Effect of online journal club on Evidence-based practice knowledge, intent, and utilization in school nurses. *Worldviews on Evidence-Based Nursing*, 9, 117–126. [PubMed: 22490083]
- Tseng V (2012). The use of research in policy and practice. *Society for Research on Child Development Social Policy Report*, 26, 1–23.
- Weiss CH (1979). The many meanings of research utilization. *Public administration Review*, 39, 426–431.
- West RF & Rhoton C (1994). School district administrators' perceptions of educational research and barriers to research utilization. *ERS Spectrum*, 12, 23–30.
- William T. Grant Foundation. (2016). *Improving the use of research evidence: An updated statement of research interests and applicant guide*. William T. Grant Foundation: New York.
- Williams D, & Coles L (2007a). Teachers' approaches to finding and using research evidence: an information literacy perspective. *Educational research*, 49, 185–206.
- Williams D & Coles L (2007b). Evidence-based practice in teaching: an information perspective. *Journal of Documentation*, 63, 812–853.
- Zaboski BA, Schrack AP, Joyce-beaulieu D, & Macinnes JW (2017). Broadening Our Understanding of Evidence-Based Practice: Effective and Discredited Interventions. *Contemporary School Psychology*, 21, 287–297.

Figure 1.
Article and instrument search and selection process

Table 1.

Instrument characteristics

#	Article(s)	Instrument Name	UC	EC	I & P	Weiss	Participants	Country
1	Adams (2007, 2009); Adams & Barron (2009, 2010)	School Nurse Evidence Based Practice (SN-EBP)	X	X	X		School nurses	United States
2	Brown & Zhang (2016a, 2016b, 2017)		X	X			Teachers	NR
3	Cahill et al. (2013)	Adapted Fresno Test	X				Occupational therapists	United States
4	Cahill et al. (2013)		X		X		Occupational therapists	United States
5	Cooper & Levin (2010, 2013); Levin et al. (2011)		X	X	X		Superintendent, (Vice) Principals, school district leaders	Canada
6	Demski & Racherbaumer (2015)		X	X	X		Principals	Germany
7	Finnigan, et al. (2013)		X	X	X		School district leaders, educators	United States
8	Hemsley-Brown & Oplatka (2005)	Barriers Scale	X	X	X		Elementary & Secondary principals and teachers	England Israel
9	Lysenko et al. (2014, 2015)	Questionnaire about the use of research-based information (QURBI)	X	X	X	X	Teachers, Administrators, Professionals	Canada
10	Massell et al. (2017)			X	X		School improvement staff, State education agency professionals	United States
11	McKee et al. (1987)		X				School psychologists	NR
12	Meline & Paradiso (2003)		X	X	X		Speech language pathologists	United States
13	Penuel et al. (2016, 2017)	Survey of Practitioners' Use of Research (SPUR)	X	X	X	X	Superintendents, principals, other senior staff	United States
14	Sortedahl (2012a)		X		X		School nurses	United States
15	Sortedahl (2012b)		X		X		School nurses	United States
16	West & Rhoton (1994)				X		Superintendents, principals, other senior staff	United States
17	Williams & Coles (2007a, b)		X	X	X		Teachers, Head teachers, School librarians, EA advisors	Scotland England Wales
18	Zaboski et al. (2017)		X				School psychologists	United States

I & P = implementation and practice, UC = user characteristics, EC = environmental characteristics

Table 2.

Quantitative Properties of the Instruments

	Scale, Subscale Reliability (α)	Validity: Evidence	Sample Size	Response Rate (%)
1	NR, 0.62–0.89	Content: Expert review of the survey Convergent: With education & professional association membership Structure: PCA	247–386	NR, 53.80%, 56.80
2	NR	Face & Construct: Piloted with teachers from the primary sector	696	65
3	NR	Content: Claimed, but with no evidence	29	NR
4	NR	NR	29	NR
5	NR	NR	188	53.71
6	NR	NR	297	NR
7	NR	NR	286	NR
8	0.82, NR	NR	105	40.6 (from England)
9	0.94, 0.77–0.92	Content: Focus groups with teachers Structure: PCA Convergent: With attitudes & experience	2425 1153	58.7 NR
10	NR	NR	~300	65–81
11	NR	NR	210	25
12	NR	NR	27	21
13	NR, .67–.93	Content: Interviews with education leaders Structure: EFA Discriminant: IRT, but not reported	733	51.5
14	NR	Face & Content: Reviewed by nurses with masters and doctoral degrees	2–13	40–100
15	NR	Face & Content: Reviewed by nurses with masters and doctoral degrees	11	31.40
16	NR	NR	543	NR
17	NR	NR	390–549	11.14–14.08
18	NR, 0.531–0.811	NR	80	26

NR = Not-reported, PCA = Principal Components Analysis, EFA = Exploratory Factor Analysis, IRT = Item Response Theory