



School educators' engagement with research: an Australian Rasch validation study

Joanne Gleeson¹ · Blake Cutler¹ · Mark Rickinson¹ · Lucas Walsh¹ · John Ehrich² · Connie Cirkony³ · Mandy Salisbury¹

Received: 27 October 2021 / Accepted: 2 February 2023 / Published online: 24 February 2023
© The Author(s) 2023, corrected publication 2023

Abstract

There are growing expectations in Australia and internationally that school educators will engage with and use research to improve their practice. In order to support educators to respond to such expectations, there is a need to be able to accurately assess the levels of educators' research engagement. At present, however, few psychometrically sound instruments are available. Drawing on two studies of Australian educators ($n = 1,311$) and utilising Rasch analysis, supported by confirmatory factor analysis, this paper reports on the development of a brief eight-item scale that demonstrates validity and reliability and evidenced unidimensionality in the second study. The scale is intended as a quick, easy-to-use tool for educators to gain insights into their beliefs about the value of engaging with research, their actions around using research, and their confidence in finding, interpreting, and judging the quality of relevant research. Notwithstanding the need for further testing, this paper argues that the scale has the potential to be applicable to other educational contexts and to contribute to future research into educators' research engagement and its assessment. The scale can also provide school and education system leaders, as well as evaluators and researchers, with data regarding educators' research engagement over time, allowing for research use supports and resources to be better targeted.

Keywords Research engagement · Research use · School educators · Scale development · Rasch analysis · Assessment

1 Introduction

Internationally, the expectation for school educators to use research to improve their practice is growing (e.g., Malin et al., 2020; Nelson & Campbell, 2019; Organisation for Economic Co-operation and Development [OECD], 2022).

✉ Joanne Gleeson
jo.gleeson@monash.edu

Extended author information available on the last page of the article

There is increasing recognition that “research-rich school and college environments are the hallmark of high performing education systems” (BERA, 2014, p. 6). Similarly, being a “good teacher” is often associated with using research alongside existing knowledge and experiences (e.g., Sachs, 2016; Wieser, 2018; Winch, 2017). As such, educators are being encouraged to be more research “literate” or “engaged”, with substantial attention having been paid to the different ways in which this can be developed (e.g., Cordingley, 2015; Evans et al., 2017).

Within Australia though, there is surprisingly little knowledge of educators’ research engagement. While some significant studies were undertaken several years ago (e.g., Biddle & Saha, 2005; Figgis et al., 2000), it has not been until very recently that a number of empirical studies on the use of research in Australian schools have emerged (e.g., Mills et al., 2021; Parker et al., 2020; Prendergast & Rickinson, 2019; Rickinson, Gleeson, Walsh, Cutler, et al., 2021; Rickinson, Gleeson, Walsh, Salisbury, et al., 2021; Rickinson et al., 2022). This situation is in contrast to other parts of the world, where research investigation on this topic has remained more rather than less prominent over time (e.g., Brown, 2015; Cain, 2019; Finnigan & Daly, 2014; Gorard, 2020). Greater insights into educational research use are therefore needed in the Australian context in order to be able to develop appropriate supports and resources for educators.

One area of particular need in Australia, but also internationally, is the development of instruments with which to accurately assess educators’ levels of research engagement. A recent systematic review of research use measures for K-12 school education found that while several instruments are available, few provide evidence of reliability or validity, and those that do “were developed and tested in a North American context and may not be suitable for measuring use of research evidence in other places” (Lawlor et al., 2019, p. 226). This finding echoes an earlier review, which noted how “psychometric inconsistencies such as lack of reliability coefficients, and small unrepresentative or misrepresentative samples are among the main limitations of the scales developed to measure use” (Dagenais et al., 2012, p. 303). This lack of psychometrically sound instruments is an impediment, not only to helping educators identify and address gaps in their research-related skills and attitudes, but also to deeper understandings of the interrelationships between research use and educators’ professional knowledge and their effective practice.

The aim of this paper is to address these important gaps in the research use literature by reporting on the development and validation of a brief research engagement scale for use with educators. Drawing on two studies of Australian educators ($n = 1,311$) and utilising Rasch analysis, with support of confirmatory factor analysis, evidence is provided of the scale’s reliability and validity, with unidimensionality shown in the second study. The work presented in this paper is part of an ongoing 5-year study (Monash Q Project) to understand and improve the use of research in Australian schools (e.g., Rickinson, Cirkony, Walsh, et al., 2021; Rickinson, Gleeson, Walsh, Salisbury, et al., 2021; Rickinson et al., 2022).

The paper has six main sections following this opening section. The next two sections briefly review the meanings of educators’ research engagement in the literature and how its measurement has been approached. There are then two sections that explain the methods and the findings of two sequential studies to test, and then re-test, the validation of the scale. The paper concludes with a discussion of the

findings of the studies, followed by a consideration of their limitations and implications for future research.

2 Educators' research engagement

Alongside increasing international expectations that educators will use research to inform and improve their practice is a growing interest in the extent to which educators are research-engaged themselves (e.g., Hill, 2022). This interest is framed by educators' enduring low engagement with research in both Australian (e.g., Parker et al., 2020; Prendergast & Rickinson, 2019; White et al., 2018, 2021) and international (e.g., Cain, 2019; Coldwell et al., 2017; OECD, 2022; Penuel et al., 2017; Walker et al., 2019) contexts, despite increasing efforts of jurisdictions worldwide that are focused on developing national and/or federal infrastructures for research use in education (Malin et al., 2020). The meaning of "research engagement", though, is variously understood (Nelson et al., 2017). In the literature, research engagement has been taken to mean both engaging in (i.e., doing) research and engaging with (i.e., reading and using) research (e.g., Borg, 2007; Prendergast & Rickinson, 2019). It has also been related to educators' research use skills and abilities (e.g., Evans et al., 2017); the different ways in which research is used, such as for conceptual or instrumental purposes (e.g., Cain, 2015); the value placed on research use (e.g., Stoll et al., 2018); and "orientations" towards research that encompass individuals' research-related attitudes, beliefs, confidence, and practices (e.g., Mills et al., 2021). Most notably, a large-scale UK-based series of studies by the National Foundation for Educational Research (NFER) (Nelson et al., 2017; Poet et al., 2015) developed a number of constructs that, when combined, created a "picture of research engagement" (Nelson et al., 2017, p. 5). These constructs covered educators' beliefs in the value of research use, knowledge about how and where to access research, as well as translation of such knowledge into action, and abilities to interpret, question, and apply research in practice.

Our conception of educators' research engagement came about through our broader work in the Monash Q Project, a 5-year initiative focused specifically on understanding and improving the use of academic research in Australian schools. Guided by NFER's work in particular (Nelson et al., 2017; Poet et al., 2015), we conceived an educator being "research engaged" if they have combined positive beliefs in the value of research use, motivation to use research in practice, confidence in their research use-related skills and capacities, and regular use of research in practice. Through investigating educators' research engagement, our intention was to improve our knowledge of those individual factors that influenced research use in practice, which in turn, would help to inform tailored system and school-level interventions to improve educators' research use.

Theoretically, our construct of research engagement was informed by our work with project partner, BehaviourWorks Australia. To understand educators' research use behaviours and the key levers of behavioural change, Michie et al.'s (2011) "COM-B" model was used as a part of this work to analyse the interactions between individuals' research use-related capabilities, opportunities, and motivations.

Findings from behavioural interviews with educators suggested that the behaviours associated with using research well in practice were driven by educators' beliefs in the value of research use and their intentions to use research, which in turn, were driven, to a large extent, by their research use-related skills, knowledge, and confidence (Plant et al., 2022).

Empirically, our construct was informed by previous studies that highlighted the important influence of individual factors on educators' engagement with research (e.g., Dagenais et al., 2012; van Schaik et al., 2018). While there are substantial bodies of work focused on school contextual factors, such as the leadership and organisational features of "research-engaged schools" (e.g., Cooper & Levin, 2013; Godfrey & Brown, 2019), or system contextual factors, such as "mediation" mechanisms that include funded dissemination of research (e.g., Cain, 2017; Tripney et al., 2014), individuals' research use beliefs, motivations, and confidence have been found as very relevant influences (van Schaik et al., 2018) and noted as pre-conditions for educators' research use in practice (Judkins et al., 2014; Lysenko et al., 2014; Penuel et al., 2017; Williams & Coles, 2007).

In our own work, we found significant associations between educators' greater use of research in practice and their positive beliefs in the value of research use, high confidence in their research-use-related capabilities, and their strong motivations to initiate research-use-related actions (e.g., initiating research discussions with colleagues) (Rickinson, Gleeson, Walsh, Cutler, et al., 2021). Additionally, qualitative findings from educators' survey and interview responses reinforced the importance of educators' research use "mindsets" and confidence as enablers of their research use (Rickinson, Gleeson, Walsh, Salisbury, et al., 2021). From a different perspective, other influential studies included those that reported on educators' low or negative beliefs and motivations regarding research use and their actual use in practice. For example, those educators who were sceptical about research value or its quality (e.g., Buske & Zlatin-Troitschanskaia, 2018), uncertain of its transferability to their specific contexts (e.g., Joram et al., 2020), not confident in their research use skills (e.g., Coldwell et al., 2017), not interested in using research (e.g., Borg & Alshumaimeri, 2012; Martinovic et al., 2012), or did not believe that research was relevant to or usable in their practice (e.g., Broekkamp & Van Hout-Wolters, 2007; Vanderlinde & van Braak, 2010) were found to use research less in practice.

3 Measures of educators' research engagement

A review of the research use literature reveals few reliable measures of educators' research engagement. Some years ago, Dagenais et al. (2012) noted a dearth of available psychometrically sound measures of educators' research use. Little appears to have changed since that time, with a recent systematic review of research use instruments in education noting that of the 18 included quantitative instruments, only a small minority provided evidence of validity (38.8%) or reliability (27.7%) (Lawlor et al., 2019, p. 225). As a result, the review team only recommended three measures as suitable for future work. However, these particular instruments were not well suited for our work because they measured a broad mix of research use influences

(e.g., individual user, environment, school context, and implementation characteristics), rather than just individuals' research-related attitudes, beliefs, and confidence levels alone. Furthermore, the other instruments that did focus on individual characteristics such as research-related perceptions and attitudes (e.g., Lysenko et al., 2014), capacity for use (e.g., Cahill et al., 2015), and user confidence (e.g., Williams & Coles, 2007) were all reported as not reliable.

Beyond those identified within Lawlor et al.'s (2019) systematic review, there are several other instruments or tools that are available, but all have limitations with respect to measuring educators' research engagement. For example, NFER's five distinct measures of research engagement, comprising 15 items in total, are not validated as an overall scale (Poet et al., 2015, p. 62). A sixth two-item construct that captures educators' knowledge of research findings and methods was also developed separately (Poet et al., 2015, p. 64). The reliability of the six measures is variable, with only two of the constructs reporting Cronbach's alpha estimates greater than an expected .70 (Poet et al., 2015, p. 16). As another example, while Stoll et al. (2018) developed a tool for educators to self-assess their "engagement with ... research evidence" (p. 3), the tool is not validated as a scale, nor have the tool's components been analysed for reliability. The tool comprises 16 indicators across three components that span: educators' understanding of research, and knowledge of how to access, assess and relate it to practice; their beliefs in the importance of research use; and their conversations about its application, as well as the degree to which research is used.

This lack of existing reliable measures of educators' research engagement highlights an important need to develop and validate new scales. As explained in the previous section, the design of the scale reported in this paper focuses on the individual characteristics of educators' research use, that is, educators' beliefs in the value of research use, the extent to which they consult research in practice, initiation of discussions and actions regarding research use, and confidence in their research use-related skills and abilities. The design of the scale aims to be brief, practical, and easily interpretable, so as not to overtax educators from a time perspective (Neal et al., 2019), and has considered its potential applicability to other education contexts (Gitomer & Crouse, 2019). Its composition has also taken account of effective professional learning design features, such as creating opportunities for educators to actively learn and reflect (Cirkony et al., 2021).

The next sections of this paper outline the methods and findings of two sequential studies to test, and then re-test, validation of the scale. In Study 1, nine initially hypothesised research engagement Likert-style rating items were included in a larger survey of 492 Australian educators about their perceptions and use of research in practice. Using the survey responses of these educators, Study 1 examined the psychometric functioning of these items as a potential scale. After eliminating one item for misfit, an eight-item scale was developed that showed evidence of validity and reliability, but was multi-dimensional. In Study 2, the eight validated items from Study 1 were included in a new survey of 819 Australian educators about their attitudes and behaviours to sharing research. Using the survey responses of this second group of educators, Study 2 was then able to re-test the eight-item scale for evidence of validity, reliability, and unidimensionality.

4 Study 1: methods for testing scale validation

4.1 Participants

An overall sample for this study was recruited from two sources (Sample A and Sample B). Sample A included nominated educators (leader/teacher/other staff) from 78 volunteer partner schools participating in the Q Project across four Australian states: New South Wales (NSW), Victoria (VIC), South Australia (SA), and Queensland (QLD). Of the 182 survey invitations sent, 125 surveys were completed (68.7% response rate). For Sample B, an external data collection agency was engaged to recruit additional research participants through their panels to both increase and diversify the overall survey respondent sample. The same online survey was administered and completed by 367 educators from the four participating states. Overall, 492 respondents from 414 schools completed the survey in Study 1 (see Table 1), and the sample demographics were generally reflective of the broader Australian teaching profession (Thomson & Hillman, 2019). It is worth noting that the high number of Victorian respondents was potentially influenced by Monash University's location being in this state.

4.2 Item development and instrument

Nine items were initially hypothesised as comprising educators' research engagement. These nine items, shown in Table 2, were situated in two quantitative questions of the larger survey (i.e., eight quantitative and eight open-text questions in total) that was administered to Australian educators. As explained in previous sections, the focus of these nine items was informed by Michie et al.'s (2011) COM-B model of behaviour, our partner work with BehaviourWorks Australia, our own empirical studies, and previous research emphasising the importance of individual characteristics as influences of educators' research use. The original development and wording of the items were carried out by three researchers in the Q Project team and were initially informed by several large-scale international studies that focused on research and evidence use in educational practice (e.g., Penuel et al., 2016; Poet et al., 2015). Three items (i.e., P4Q11, "I am not confident in how to judge the quality of the research"; P4Q12, "I am not clear about how research can be used to help change practice"; and P4Q18, "I don't believe research will help to improve student outcomes") were negatively worded to help prevent and detect acquiescence response bias (i.e., where respondents agree with all items indiscriminately), with these items being reverse-coded during analysis (Rattray & Jones, 2007). One item (P4Q13, "I believe teacher observations and experience should be prioritised over external research") was positively worded, however, was intended to capture educators' preferences for or intentions to use their own knowledge and experience relative to academic research, and was reverse-coded during analysis. All items were tested with a number of Australian educators and policy-makers, as well as international and Australian advisors to the Q Project, as part of a broader, four-wave

Table 1 Study 1 and Study 2 samples (combined $n = 1,311$)

Study	1		2	
	<i>n</i>	%	<i>n</i>	%
Overall sample size	492		819	
School sector ^a				
Government or public	325	66.1	564	68.9
Independent or private	77	15.7	122	14.9
Catholic	90	18.3	124	15.1
School type				
Primary (K-6)	205	41.7	459	56.0
Secondary (7-12)	156	31.7	253	30.9
Combined (K-12)	117	23.8	107	13.1
Special	14	2.8	--	--
Self-identified gender ^b				
Woman	372	75.6	727	88.8
Man	118	24.0	91	11.1
Self-entered option	2	0.4	--	--
Location				
Victoria (VIC)	195	39.6	228	27.8
New South Wales (NSW)	149	30.3	225	27.5
Queensland (QLD)	116	23.6	174	21.2
South Australia (SA)	32	6.6	62	7.6
Western Australia (WA)	--	--	87	10.6
Tasmania (TAS)	--	--	27	3.3
Australian Capital Territory (ACT)	--	--	10	1.2
Northern Territory (NT)	--	--	6	0.6
Role				
Senior leader	99	20.1	39	4.8
Middle leader	60	12.2	60	7.3
Teacher	281	57.1	589	71.9
Other staff (e.g., librarian)	52	10.6	131	16.0
Qualification ^c				
Bachelors	273	55.5	544	66.4
Postgraduate	219	44.5	246	30.0
Experience in years ^d				
0 to < 5	74	15.0	232	28.3
5 to < 10	76	15.4	243	29.7
10 or more	341	69.3	344	42.0

^aNine participants in Study 2 did not provide their school sector

^bOne participant in Study 2 did not disclose their gender

^c29 participants in Study 2 did not disclose their qualification level

^dOne participant in both Study 1 and Study 2 did not provide their years of experience

Table 2 Nine hypothesised research-engagement items

Item ID	Item statement	Item scale
Lead-in instruction 1		
<i>Thinking generally about school initiatives that you have implemented in the last 12 months, how often did you consult the following information sources to help inform your decision?</i>		
P2Q24	Articles, reports, books or summaries based on the work of universities or research organisations	5-point Likert: <i>Never, Rarely, Sometimes, Often, Always</i>
P2Q2L	Advice or guidance based on research from a university or research organisation	
Lead-in instruction 2		
<i>Please rate the extent to which you agree or disagree with the following statements about your own awareness/use of research.</i>		
P4Q11	I am not confident in how to judge the quality of research*	5-point Likert: <i>Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree</i>
P4Q12	I am not clear about how research can be used to help change practice*	
P4Q13	I believe teacher observations and experience should be prioritised over external research*	
P4Q16	I regularly initiate discussions regarding research and its connection with school practice	
P4Q18	I don't believe research will help to improve student outcomes*	
P4Q19	When confronted with a new problem or decision, I look for research that might be relevant	
P4Q1L	I feel confident in analysing and interpreting research for my own teaching context	

*Indicates items which were reverse-coded during the analysis

piloting approach that was led by an external research consultancy to the refine overall survey design and item construction (for details about survey design and testing, see Rickinson, Gleeson, Walsh, Cutler, et al., 2021).

4.3 Procedure and data analysis

During March to August 2020, the nominated partner school educators were emailed a personalised, identifiable link to a Monash-licensed Qualtrics online survey. Each survey was expected to take approximately 20 minutes to complete. During August to September 2020, the external agency administered the same survey using their own software platform to their recruited participants to protect anonymity, but included additional demographic questions (e.g., school name) to enable school profile data (e.g., location) to be sourced from the Australian Curriculum, Assessment, and Reporting Authority website for each respondent. The external agency then provided anonymised quantitative and qualitative data to the Q Project team in spreadsheets for analysis.

To prepare the dataset for this study's analysis, the nine Likert-style rating items were assigned numeric ratings of 1 to 5 (either strongly disagree, disagree, neutral, agree, strongly agree; or never, rarely, sometimes, often, always; reverse-coding considered) using IBM SPSS Statistics software (Version 27.0). To benefit scale validation, combined Rasch (1960) and classical test theory (CTT) analytical approaches were used to examine the validity, reliability, and unidimensionality of a potential scale (Curtis, 2005). Firstly, an initial CFA was conducted in R (R Core Team, 2021) with the lavaan package (Version 0.6-9; Rosseel, 2012) to determine whether all of the items could be loaded onto one component (Curtis, 2005). Diagonally weighted least squares estimation (DWLS) was selected as the estimation method because it was considered appropriate for ordinal data and provided less biased factor loading estimates than other estimation methods (Li, 2016). Regression weights of each residual to each observed variable were set to 1.0, and the variance for the latent variable was set to 1.0 (variance standardisation method). Fit statistics used were chi-square (χ^2 , significant p value is acceptable; Schermelleh-Engel et al., 2003), normed chi-square (χ^2/df , < 5.0 suggests good fit; Schumacker & Lomax, 2004), comparative fit index (CFI, $\geq .95$ expected); goodness of fit index (GFI, $\geq .95$ expected), standardised root mean square residual (SRMR, $\leq .08$ expected), and the root mean square error of approximation (RMSEA, $\leq .08$ expected) (Hooper et al., 2008; Schreiber et al., 2006).

Secondly, Rasch analysis, through its model requirements of invariance and local independence, was considered a more appropriate approach to examine evidence of scale unidimensionality, as well as to perform item selection and reduction (Andrich, 2011; Tennant et al., 2004). Analysis using the polytomous Rasch model with partial credit parameterization (Masters, 1982) was conducted using RUMM2030 software (Andrich et al., 2013). Given the utilisation of 5-point Likert-style rating items, a polytomous model allowed for more than two possible response categories. The partial credit model, with its ability to allow the distances between response categories to emerge from the data and be non-uniform, accommodated the

assumption of likely different item interpretations across respondent types (Masters, 1982). Steps for conducting Rasch analyses have been described in detail elsewhere (e.g., Pallant & Tennant, 2007; Tennant & Conaghan, 2007; Tennant et al., 2004) and are not repeated here. These resources provide procedures for assessing model fit (item–person interaction, item–trait interaction, individual item, and person fit), internal reliability (person separation index (PSI)), suitability of response options (response category thresholds), item bias (differential item functioning (DIF)), inter-item correlations (response dependency), and dimensionality. The structure of this paper has been guided by other examples where the detailed steps for conducting Rasch analyses have not been included (e.g., Creed et al., 2016).

Finally, another CFA was conducted, again using the lavaan package in R with DWLS estimation to confirm component structure and test for convergent validity. Convergent validity was examined using item loadings at (or near) $\geq .50$ (Widaman, 1985). Additional reliability measures of Cronbach's alpha (α) and McDonald's omega (ω) were also calculated in SPSS at the end of each study, with $> .70$ being acceptable for research purposes (Hayes & Coutts, 2020; Nunnally & Bernstein, 1994).

5 Study 1: results from testing scale validation

5.1 Initial CFA

An initial CFA was conducted where all nine items were allowed to load onto a single latent variable to test the a priori assumption that all items constituted a unidimensional measure of research engagement. One pair of residuals were covaried, between the two items with a unique response scale (i.e., P2Q24-P2Q2L), reflecting the possible influence of a methods factor. The model demonstrated reasonable fit, with only the CFI being marginally lower than the expected cut-off: $\chi^2 = 104.822$, $df = 26$, and $p < .001$; $\chi^2/df = 4.031$; GFI = .972; CFI = .937; SRMR = .075; RMSEA = .079. The standardised path coefficients were all statistically significant at $p < .001$, with all item loadings above .40: P2Q24 = .555; P2Q2L = .470; P4Q11 = .403; P4Q12 = .558; P4Q13 = .426; P4Q16 = .509; P4Q18 = .608; P4Q19 = .592; P4Q1L = .490. The results suggested that the nine items could reasonably be represented as a single latent factor and be examined as such using the Rasch model.

5.2 Rasch analysis

5.2.1 Initial model fit

Using the same nine items, an initial inspection of the summary statistics suggested that the data did not conform to the Rasch model. The item fit residual mean and standard deviation (*SD*) were 0.56 and 2.01 respectively. This *SD* was well above the theoretical value of 1 suggesting that overall, the pattern of responses across items was not as expected. The person fit residual mean of 0.49, with an *SD* of 0.98, while

suggesting a slight underfit with some inconsistency in some response patterns, indicated that the overall response profile across persons was more or less as expected. Consistent with these results, the item-trait interaction was significant at $x^2 = 87.50$, $df = 63$, and $p = .022$, suggesting that the hierarchical ordering of items was not consistent across all levels of the underlying trait.

No persons showed individual fit statistics outside of the acceptable ± 2.5 range; hence, none were removed at this stage. Two items showed misfit to model expectations with fit residual values $> \pm 2.5$. These included P4Q11 (“confident to judge”, fit residual = 2.70) and P4Q13 (“prioritise teacher experience over research”, fit residual = 4.54).

5.2.2 Thresholds

A “good fitting” model is indicated by an ordered set of response thresholds for each item (Tennant & Conaghan, 2007). Upon examination of the item threshold map, four items displayed disordered thresholds, of which one was the misfitting item, P4Q11. Each of these items’ category probability curves displayed poor discrimination at the thresholds between categories 1 and 2 or 2 and 3. One way of overcoming disordered thresholds is to rescore items by collapsing relevant adjacent response categories (Andrich & Marais, 2019; Wetzel & Carstensen, 2014). Rescoring the four misfitting items resulted in all nine items then showing ordered thresholds, with the fit residual value of P4Q11 (1.56) resolved within the acceptable ± 2.5 range. However, the overall model fit was not improved, with x^2 increasing to 92.04, $df = 63$, and $p = .009$, necessitating further item resolution actions.

5.2.3 Differential item functioning

Differential item functioning (DIF) occurs when items do not function in the same way for different groups of people, who otherwise have the same value on the trait (Andrich & Marais, 2019). In this study, DIF was assessed by respondents’ role (i.e., teacher, leader, or staff in other roles, such as librarians), gender (i.e., man, woman, or self-described), years of experience (i.e., less than 5 years, 5–10 years, 10–15 years, or 15+ years), and/or qualifications held (i.e., undergraduate degree, such as a Bachelor degree, post-graduate degree by coursework, or post-graduate degree by research). Previous analysis reported by the Monash Q Project indicated differences in response patterns for each of the nine items by these person factors (Rickinson, Gleeson, Walsh, Cutler, et al., 2021).

Analysis of variance of the residuals indicated that one item, P4Q18 (“link to improved student outcomes”), showed statistically significant uniform DIF (i.e., where differences are uniform across all levels of the trait) by role, with a mean square value (MS) of 6.966, $F = 9.319$, $p < .001$. The item also showed non-uniform DIF (i.e., when a difference occurs at some but not all levels of a trait) by role, gender, experience, and qualification. Examination of the item characteristic curve (ICC) showed slight over-discrimination (i.e., observed means steeper than the curve). The item was split by role, after which all item splits showed fit residual

values within the acceptable ± 2.5 range and were retained in the model. This action also resolved the item's non-uniform DIF. No other DIF was detected.

5.2.4 Item fit

After testing for DIF, P4Q13 (“prioritise teacher experience over research”) still did not fit the model with a fit residual of 3.46. Examination of the ICC showed slight under-discrimination (i.e., observed means flatter than the curve). The item was removed from the model.

5.2.5 Response dependency

An initial inspection of the item residual correlation matrix provided only one incidence of response dependence, that is, where the response to one item is governed by the response to the previous item (Marais & Andrich, 2008; Smith, 2002). Specifically, a high correlation of standardised residuals (i.e., $r > +0.3$) between pairs of items is problematic and suggests that the items are more dependent in their responses than can be accounted for by the persons' and items' locations. Items P2Q24 (“university research”) and P2Q2L (“university guidance”) showed an initial residual correlation of $r = .376$, which decreased after item resolution actions were taken ($r = .297$). No other response dependency was detected.

5.2.6 Revised model fit

A final inspection of the summary statistics, based on the remaining eight items, suggested that the data conformed to the Rasch model. The item fit residual mean and *SD* were 0.17 and 1.03 respectively. The person fit residual mean was 1.04 with an *SD* of 1.13. Consistent with these results, χ^2 improved to 82.652, $df = 70$, and $p = .142$. No persons showed individual fit statistics outside of the acceptable ± 2.5 range, nor did any individual items. The power of analysis fit was reported as “good”.

5.2.7 Person separation and scale reliability measures

The PSI statistic in the Rasch model is an indicator of how well a group of items is able to differentiate respondents with different levels of the trait. It can be interpreted similarly to Cronbach's alpha, with $> .70$ being acceptable for research purposes (Nunnally & Bernstein, 1994). The original PSI value, both with and without extreme scores, was .79. The original Cronbach's alpha, both with and without extreme scores, was .77. After item resolution, the eight-item scale demonstrated acceptable reliability with a PSI value of .78 (without extremes), Cronbach's alpha value of .76, and McDonald's omega value of .77.

5.2.8 Targeting of the revised scale

A final inspection of the person-item threshold distribution, including the average mean personal location value of 1.04 ($SD = 1.13$), indicated that the scale was reasonably well targeted, with the items overall appearing to be spread across a relatively wide range of difficulty. Overall, respondents scored above the average of the scale mean (i.e., zero), with key differences noted between school leaders (mean = 1.72, $SD = 1.19$) when compared with teachers (mean = 0.75, $SD = 0.94$), and educators holding undergraduate qualifications (mean = 0.79, $SD = 1.00$) when compared with those holding postgraduate coursework (mean = 1.25, $SD = 1.20$) and research qualifications (mean = 1.96, $SD = 1.15$). These differences were not unexpected, given analysis previously reported by the research team (Rickinson, Gleeson, Walsh, Cutler, et al., 2021).

5.2.9 Unidimensionality

The results of an initial PCA of the standardised residuals indicated that the residuals were patterned onto more than one subscale. Specifically, items associated with consulting research (i.e., P2Q24, “university research”; and P2Q2L, “university guidance”) and undertaking research-use actions (i.e., P4Q16, “initiate discussions”; and P4Q19, “look for research”) loaded positively onto the first principal component (termed “research-use and action” component), while the remaining items, those associated with “research-related beliefs and confidence” (i.e., P4Q11, “confident to judge”; P4Q12, “link to changed practice”; P4Q18, “link to improved student outcomes”; and P4Q1L, “confident to analyse”), loaded negatively onto the same component. A subtest analysis was performed to examine the degree of multidimensionality, with items grouped into two subtests informed by the PCA results and logical sense (i.e., items grouped by similar type) (Marais & Andrich, 2008). The four positive-loading “research-use and action” items were combined as one higher-order polytomous item, with three of the negative-loading “research-related beliefs and confidence items” used as another. It was decided not to combine the splits of the original item, P4Q18 (“link to improved student outcomes”), because of smaller response sizes. Testing via a paired t test, showed that multidimensionality was evident, with twenty-nine cases (out of 492, including extremes) of difference in person estimates identified (5.89%) at the 95% confidence level and five (1.02%) at the 99% confidence level. This result only just exceeded the critical value of 5% and was considered a minor violation only of independence (Andrich & Marais, 2019; Smith, 2002).

However, multidimensionality was evident when comparing subtest analysis to original reliability estimates for inflation. The PSI without extremes of .69 obtained via subtest analysis was lower than the original PSI without extremes of .78, suggesting an inflated reliability due to trait dependence (Andrich, 2016). Following subtest analysis, overall model fit improved, with $\chi^2 = 41.166$, $df = 35$, and $p = .218$.

5.3 Final CFA

To confirm this single-component, eight-item scale, a final CFA, using diagonally weighted least squares estimation, was undertaken. Two pairs of error residuals were covaried (P2Q24-P2Q2L and P4Q11-P4Q19). As noted earlier, the first covariance reflects the possible influence of a methods factor, while the covariance between P4Q11 and P4Q19 is considered appropriate given the connection between educators' research use confidence and how regularly they source research is well supported by our own research findings (Rickinson, Gleeson, Walsh, Cutler, et al., 2021), as well as various international studies (e.g., Judkins et al., 2014; Lysenko et al., 2014; Penuel et al., 2017; Williams & Coles, 2007). Fit of a single-component model was as follows: $\chi^2 = 65.623$, $df = 18$, and $p < .001$; $\chi^2/df = 3.646$; GFI = .980; CFI = .954; SRMR = .068; and RMSEA = .073. The standardised path coefficients were all statistically significant at $p < .001$, with all item loadings above .40: P2Q24 = .531; P2Q2L = .436; P4Q11 = .445; P4Q12 = .547; P4Q16 = .527; P4Q18 = .556; P4Q19 = .659; P4Q1L = .536.

6 Study 1: initial scale validation conclusion

Overall, Rasch and CFA analysis approaches were used to examine the psychometric functioning of a potential short-form research engagement scale. An eight-item scale ("research-use and action" items: P2Q24, P2Q2L, P4Q16, P4Q19; and "research-related beliefs and confidence" items: P4Q11, P4Q12, P4Q18, P4Q1L) was developed that showed evidence of validity and reliability. While a final CFA confirmed a single-component scale and the results of a paired t test indicated only a very minor violation of independence, further Rasch testing indicated inflated reliability estimates of the original model when compared with the subtest analysis. The aim of Study 2 was to re-test this scale and in particular, its unidimensionality.

7 Study 2: methods for re-testing scale validation

7.1 Participants

The sample for Study 2 was recruited by an external research consultancy who were engaged to help design and then administer a new survey focused on exploring Australian educators' attitudes and behaviours to sharing educational research. Of the 849 surveys completed, 819 valid responses were included in the sample for this study (see Table 1). Given the increased representation of primary schools in the Study 2 sample as compared with Study 1, the higher proportion of women in the Study 2 sample is more reflective of the broader Australian teaching profession (Thomson & Hillman, 2019).

7.2 Instrument

The new survey comprised 51 quantitative and two qualitative questions in total. The survey was expected to take between 20 and 30 minutes and was structured such that respondents were asked to complete 20 core questions, with 33 follow-on questions proposed depending on responses to core questions. The eight items from Study 1 were included in one of the core survey questions and were presented in the same format and using the same wording.

7.3 Procedure and data analysis

During May to July 2021, the external research consultancy administered the online survey using their own software platform to their recruited participants to protect anonymity. All anonymised quantitative data was provided by the consultancy to the research team in spreadsheets for analysis. The same software applications and procedures from Study 1 were used to conduct Rasch analysis in Study 2.

8 Study 2: results from re-testing scale validation

8.1 Rasch analysis

8.1.1 Initial model fit

Using the same validated eight items from Study 1, an initial inspection of the summary statistics suggested that the data did not conform to the Rasch model. The item fit residual mean and standard deviation (*SD*) were 0.25 and 2.54 respectively. Two people had extreme scores outside of the acceptable $> \pm 2.5$ range and were excluded from the analysis. The person fit residual mean (excluding extreme scores, $n = 817$) of 0.40 with an *SD* of 1.02 indicated that the overall response profile across persons was more or less as expected. The item-trait interaction was $\chi^2 = 154.125$, $df = 72$, and $p < .001$; the PSI value without extreme scores was .76; and Cronbach's alpha was .75.

Three items showed misfit to model expectations with fit residual values $> \pm 2.5$ including P2Q24 (“university research”, fit residual = -2.76), P4Q11 (“confident to judge”, fit residual = 4.82), and P4Q12 (“linked to changed practice”, fit residual = 2.79). Item P2Q24 was the only item to display disordered thresholds, with poor discrimination evident at the thresholds between categories 1 and 2. Rescoring this item by collapsing these response categories ordered the item's thresholds and resolved the item's fit (fit residual = -2.157). Items P4Q11 and P4Q12 still showed fit residual values $> \pm 2.5$, necessitating further item resolution actions.

8.1.2 Differential item functioning

DIF was again assessed by the person factors (i.e., role, gender, years of experience, and qualifications) used in Study 1. Initial analysis of variance of the residuals

Table 3 Study 2 differential item functioning test outcomes ($n = 819$)

Item	Issue	Action	Outcome	Subsequent action
P4Q11, fit residual = 4.60	Significant non-uniform DIF by role, $MS = 4.06, F = 3.72, p < .001$	Split by role	P4Q11T (teachers), fit residual = 3.23; subsequently resolved through other actions	P4Q11S for poor fitting ICC
P2Q24, fit residual = -2.11	Significant non-uniform DIF by role, $MS = 3.52, F = 4.74, p < .001$	Split by role	All splits showing fit residuals within acceptable ± 2.5 range	
P4Q12, fit residual = 3.86	Significant non-uniform DIF by role, $MS = 4.05, F = 3.84, p < .001$	Split by role	P4Q12T (teachers), fit residual = 3.38	Deleted P4Q12L (leaders) for disordered thresholds
P2Q2L, fit residual = -1.71	Significant non-uniform DIF by role, $MS = 2.32, F = 3.26, p < .001$	Split by role	All splits showing fit residuals within acceptable ± 2.5 range	Deleted P2Q2LS for disordered thresholds
P4Q12T (teachers), fit residual = 3.38	Significant non-uniform DIF by gender, $MS = 3.64, F = 3.45, p < .001$	Split by gender	All remaining splits showing fit residuals within acceptable ± 2.5 range	Deleted P2Q2TM (men teachers) for disordered thresholds
P2Q2LT (teachers), fit residual = -1.05	Significant uniform DIF by experience, $MS = 5.00, F = 6.10, p < .001$	Split by experience	All remaining splits showing fit residuals within acceptable ± 2.5 range	Deleted P2Q2LT1 (<5 years of experience) and P2Q2LT3 (10–15 years of experience) for disordered thresholds
P4Q11L (leaders), fit residual = 2.68	Significant uniform DIF by gender, $MS = 4.37, F = 5.64, p < .001$	Split by gender	All remaining splits showing fit residuals within acceptable ± 2.5 range	Deleted P4Q11LM (men) for disordered thresholds
P2Q2LL (leaders), fit residual = 0.51	Significant uniform DIF by gender, $MS = 3.32, F = 3.18, p < .001$	Split by gender	All remaining splits showing fit residuals within acceptable ± 2.5 range	

indicated that three items, P2Q24 (“university research”), P4Q11 (“confident to judge”), and P4Q1L (“confident to analyse”), all showed statistically significant non-uniform DIF by role, gender, and experience. Table 3 shows the steps taken, in sequential order, to resolve DIF.

As indicated in Table 3, three items, P2Q2L (“university guidance”), P4Q11 (“confident to judge”), and P4Q12 (“link to changed practice”), showed significant non-uniform and uniform DIF across more than one person factor as actions were undertaken. These items needed to be split multiple times for resolution. Other than the exclusions reported in Table 3, no other action was taken to remove items/item splits, as each time this action was taken, model fit statistics worsened.

8.1.3 Response dependency

An initial inspection of the item residual correlation matrix provided only one incidence of response dependence. Items P2Q24 (“university research”) and P2Q2L (“university guidance”) showed an initial residual correlation of $r = .460$, which was eliminated when these two items were split to resolve DIF issues. No other response dependency was detected.

8.1.4 Revised model fit

A final inspection of the summary statistics, based on all eight items, suggested that the data conformed to the Rasch model. The item fit residual mean and *SD* were 0.12 and 1.04 respectively. The person fit residual mean (without extremes) was 0.15, with an *SD* of 1.13. Consistent with these results, χ^2 improved to 120.667, $df = 103$, and $p = .112$. No items showed individual fit residual values outside of the acceptable ± 2.5 range. The power of analysis fit was reported as “good”.

8.1.5 Person separation and scale reliability measures

The scale demonstrated acceptable reliability, showing a PSI (without extremes) of .71, Cronbach’s alpha of .76, and McDonald’s omega of .75.

8.1.6 Targeting of the revised scale

A final inspection of the person-item threshold distribution indicated that the scale was reasonably well targeted, with the items overall appearing to be spread across a relatively wide range of difficulty. Overall, respondents scored above the average of the scale mean. Key differences in average mean personal location values by role and qualifications, as noted in Study 1, were not noted to the same extent in Study 2. This finding may potentially reflect the different composition of the Study 2 sample, and therefore response patterns, when compared with Study 1.

8.1.7 Unidimensionality

Similar to Study 1, “research-use and action” items loaded positively onto the first principal component, while the “research-related beliefs and confidence” items loaded negatively onto the same component. A subtest analysis was performed to examine the degree of multidimensionality, with original “research-use and action” items (i.e., P4Q16, “initiate discussions”; and P4Q19, “look for research”) combined as one higher-order polytomous item and original “research-related belief and confidence” items (i.e., P4Q18, ‘link to improved student outcomes’; and P4Q1L, ‘confident to analyse’) combined as another. It was decided not to combine the remaining items because only splits remained from the original items and subsequently had small response sizes. Testing via a paired *t* test showed that multidimensionality was evident, yet not significant enough to reject unidimensionality (Andrich, 2016). Twelve cases (out of 819, including extremes) of difference in person estimates were identified (1.47%) at the 95% confidence level, and only three (0.37%) at the 99% confidence level.

Multidimensionality was only slightly evident when comparing subtest analysis to original reliability estimates for inflation. The PSI without extremes of .70 obtained via subtest analysis was slightly lower than the original PSI without extremes of .71 suggesting a very slightly inflated reliability due to trait dependence, but not significant enough to reject unidimensionality (Andrich, 2016). Following subtest analysis, overall model fit also improved, with $\chi^2 = 95.592$, $df = 85$, and $p = .202$. The power of analysis fit was reported as “good”.

9 Study 2: re-testing scale validation conclusion

Utilising Rasch analysis, Study 2 validated a reliable and unidimensional scale comprised of the same eight items as for Study 1 (“research-use and action” items: P2Q24, P2Q2L, P4Q16, P4Q19; and “research-related beliefs and confidence” items: P4Q11, P4Q12, P4Q18, P4Q1L).

10 Discussion

There is a growing body of international evidence that suggests that when educators engage with research, their teaching skills and confidence improve (e.g., Bell et al., 2010; Godfrey, 2016), as does overall school performance (e.g., Mincu, 2014). There is also increasing emphasis on research use in schools as a means to improve educational practice in many countries (e.g., BERA, 2014; Malin et al., 2020; Nelson & Campbell, 2019; OECD, 2022). Yet, greater insights are needed into educators’ research engagement, and to this end, measurement is necessary. That there are so few valid and reliable instruments to measure this important aspect of educators’ practice represents a significant gap in the literature. This limits educators’ abilities to quickly and easily assess their own research engagement, and then to target areas of improvement that can be incorporated into individual and school performance

plans. Furthermore, it limits school and education system leaders' knowledge of educators' beliefs and attitudes towards research use, their confidence in research-related skills, and their actual use of research in practice. The lack of reliable instruments also presents real challenges for research and evaluation studies of research engagement trends over time. Without these insights, the prioritisation, development, provision, and evaluation of appropriate resources and supports for educators by universities, research organisations, policy-makers, and professional learning providers are constrained.

In response, the work reported in this paper aimed to develop an efficient measure of research engagement that demonstrates reliable and valid scores. Drawing on our own ongoing work with Australian educators, as well as findings from international studies suggesting that educators' attitudes, beliefs, and confidence levels are a condition of their research use in practice, nine items were initially hypothesised as representing educators' research engagement. Rasch analysis was used as a preferred approach to test and then re-test the scale in both studies for a number of reasons. Foremost, Rasch analysis overcomes CTT's assumption of a linear relationship between items and the latent variable by transforming ordinal scores into a linear interval-level variable. This means that Rasch analysis is particularly helpful for testing the internal construct validity of the scale for unidimensionality, assessing DIF, selecting and reducing items in the instrument through testing the invariance of the items, and ensuring appropriate ordering of the different response categories across items (Andrich, 2011; Tennant & Conaghan, 2007; Tennant et al., 2004).

In Study 1, using the survey responses of 492 educators, the nine hypothesised research engagement items were reduced to eight (P4Q13 "I believe teacher observations and experience should be prioritised over external research" was removed due to item misfit) and validated as a reliable, yet multi-dimensional scale. A final CFA confirmed a single-component scale, and the results of a paired *t* test indicated only a very minor violation of independence that would not reject unidimensionality in, and of, itself. However, scale multi-dimensionality was evident when comparing original and subtest analysis reliability estimates. As suggested by Smith (2002), multi-dimensionality is not necessarily problematic: "multidimensionality only becomes a problem when data represent two or more dimensions so disparate or distinct that it is no longer clear what dimension the Rasch model is defining or when the different subsets of items would lead to different norm or criterion-reference decisions" (p. 214). It is worth noting that the overall model fit in Study 1 improved following subtest analysis, meaning that the scale, at that stage, was measuring different, but related aspects of research engagement. Given this, sub-scores of the two dimensions, "research-use and action" and "research-related beliefs and confidence", would be used instead of a single total score for "research engagement".

In Study 2, using new survey responses of 819 educators, the eight items from Study 1 were re-tested. While the results evidenced scale validity, reliability, and unidimensionality, two issues are worth noting. Firstly, several of the items in Study 2 required improvement due to uniform and non-uniform DIF based on multiple person factors. In some cases, this meant items were split several times on different person factors to resolve item misfit, which potentially raises questions about whether the items should be retained in the model. These types of issues were not

experienced in Study 1. The necessity of these actions in Study 2 may reflect the different composition of the sample when compared with that of Study 1. In Study 2, for example, the sample included greater proportions of teachers than leaders and smaller proportions of less qualified and/or experienced educators. Elsewhere, we have reported on the different research use-related attitudes and behaviours of educators based on person factors (Rickinson, Gleeson, Walsh, Cutler, et al., 2021). Rasch analysis allowed for item improvement actions due to DIF, which enabled all original items to be retained in the scale (Andrich & Marais, 2019). This is unlike CTT methods of item reduction that rely on item-total correlations and/or indices of internal consistency, which can negatively impact the sensitivity of an instrument and its ability to provide valid scores, particularly at the extremes of the construct range (Andrich, 2011; Tennant et al., 2004). In fact, removing the problematic items significantly reduced model fit. However, to effectively investigate the influence of person factor differences and item functionality, we suggest that future research involves testing of the scale with different sample compositions and sizes.

Secondly, while Study 2 results showed evidence of unidimensionality, suggesting that the scale can be used as a higher-order single measure of “research engagement”, this outcome is different to that of Study 1. This does not mean, necessarily, that we cannot conclude that the eight-item research engagement scale shows viability as a reliable and unidimensional scale. We do suggest, though, that there is a need to conduct further testing of the scale to conclusively establish its dimensionality.

Overall, the final scale presents as a viable measure of research engagement, comprising four “research-use and action” items (P2Q24, P2Q2L, P4Q16, and P4Q19) and four “research-related beliefs and confidence” items (P4Q11, P4Q12, P4Q18, and P4Q1L), that evidences unidimensionality, reliability, and convergent validity. The scale appears to be well-targeted to educators and is presented with details of its exact wording, which, as noted by Lawlor et al. (2019) and Gitomer and Crouse (2019), is a practice that has often been missing from previous studies on this topic. The scale is also brief, easily interpreted, and quick to use (Neal et al., 2019). Single total scores obtained from the scale or sub-scores of the two dimensions, if multi-dimensional, allow educators to gain an initial and immediate insight into the extent to which they are research engaged. This will help them to identify gaps in their research-related skills and attitudes that can be actioned for improvement. These types of scores could also provide school and education system leaders, as well as evaluators and researchers, with data regarding educators’ research engagement over time, allowing for supports and resources to be targeted and provided to increase and improve research use in educational practice. Finally, it is a scale that, given its design and wording, has the potential to be applicable in other education contexts (Gitomer & Crouse, 2019).

11 Limitations and future directions

Notwithstanding the shortcomings already noted, there are several additional limitations that should be considered in relation to the future use and/or development of this scale.

Firstly, the research engagement instrument is a new scale that should continue to be tested, involving replications of these results and the use of this scale in different types of studies and contexts. While this short-form scale will provide greater insights into educators' research engagement than currently exist, future research could focus on further developments to this scale and/or the development of a longer-form version and any additional insights that may result. Furthermore, while the scale shows promise to detect change in educators' research engagement over time, it has only been tested at single points in time. An area for future research, therefore, could be the longitudinal investigation of educators' research engagement trends over time using this scale.

Secondly, while the present studies utilised a number of person factors in the analyses (i.e., role, gender, years of experience, and qualifications held), other individual factors, such as cultural background or socio-economic status, or school factors, such as the type of school (i.e., primary or secondary school), jurisdiction (i.e., government or independent school), or location (i.e., rural or metropolitan), may also influence the response patterns of individuals and, therefore, the functionality of the scale. To effectively investigate the influence of person factor differences, future research might involve the development of different scale versions and/or further developments to this scale based on different person factors. Furthermore, this scale was conducted with an Australian population. Testing its functionality, as well as its usefulness, with non-Australian populations is necessary to establish the scale's accuracy and applicability within different educational systems.

Thirdly, the scope of the present studies meant that the divergent validity of the scale has not been tested, and convergent validity has only been tested through the examination of CFA item loadings. This limitation needs to be addressed in the scope of future research (Messick, 1989).

Fourthly, the scale is based on self-report survey responses. This can be problematic as a way of measuring research engagement because of a reliance on respondent recall or the risk of a social desirability bias in responses (Gitomer & Crouse, 2019; Lawlor et al., 2019; Neal et al., 2019). An area of future research could investigate how this scale could be complemented by or used in combination with, for example, observation-based approaches to mitigate these potential issues.

Finally, the scale focuses on whether educators engage with and use research, not on *how well* they engage with or use research. As part of the ongoing work of the Q Project on "quality use of research evidence" in education, distinguishing between *increasing* and *improving* the use of research in educational practice is important (e.g., Rickinson, Gleeson, Walsh, Salisbury, et al., 2021; Rickinson et al., 2022). There is, therefore, a future opportunity to develop and validate a scale to measure educators' "quality" research use.

Author contribution All Monash University-based and University of Tasmania-based authors contributed to the study conception, design, material preparation, and data collection. Primary data analysis and original drafting of the manuscript were completed by Joanne Gleeson and Blake Cutler. Supervision of the analysis was provided by John Erich. All authors provided critical feedback on the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability The participants of this study did not give written consent for their data to be shared publicly. Supporting data is therefore not available.

Declarations

The work reported in this article is part of the Monash Q Project, which is a partnership funded by the Paul Ramsay Foundation. The authors have no other financial or non-financial competing interests to declare. This study was approved by the Monash University Human Research Ethics Committee and relevant ethics committees from each jurisdiction.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics and Outcomes Research*, 11(5), 571–585. <https://doi.org/10.1586/erp.11.59>
- Andrich, D. (2016). Components of variance of scales with a bifactor subscale structure from two calculations of α . *Educational Measurement: Issues and Practice*, 35(4), 25–30. <https://doi.org/10.1111/emip.12132>
- Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Springer.
- Andrich, D., Sheridan, B. S., & Luo, G. (2013). *RUMM2030: A MS Windows computer program for the analysis of data according to Rasch unidimensional models for measurement*. RUMM Laboratory.
- Bell, M., Cordingley, P., Isham, C., & Davis, R. (2010). *Report of professional practitioner use of research review: Practitioner engagement in and/or with research*. CUREE, GTCE, LSIS, & NTRP. <http://www.curee.co.uk/node/2303>
- Biddle, B. J., & Saha, L. J. (2005). *The untested accusation: Principals, research knowledge, and policy making in schools*. Scarecrow Education.
- Borg, S. (2007). Research engagement in English language teaching. *Teaching and Teacher Education*, 23, 731–747. <https://doi.org/10.1016/j.tate.2006.03.012>
- Borg, S., & Alshumaimeri, Y. (2012). University teacher educators' research engagement: Perspectives from Saudi Arabia. *Teaching and Teacher Education*, 28, 347–356. <https://doi.org/10.1016/j.tate.2011.10.011>
- British Educational Research Association [BERA]. (2014). *Research and the teaching profession: Building the capacity for a self-improving education system. Final report of the BERA-RSA inquiry into the role of research in teacher education*. <https://www.thersa.org/globalassets/pdfs/bera-rsa-research-teaching-profession-full-report-for-web-2.pdf>
- Broekkamp, H., & Van Hout-Wolders, B. (2007). The gap between educational research and practice: A literature review. *Educational Research and Evaluation*, 13, 220–303. <https://doi.org/10.1080/13803610701626127>
- Brown, C. (Ed.). (2015). *Leading the use of research and evidence in schools*. Institute of Education Press.
- Buske, R., & Zlatkin-Troitschanskaia, O. (2018). Investigating principals' data use in school: The impact of evidence-oriented attitudes and epistemological beliefs. *Educational Management Administration & Leadership*, 47(6), 925–942. <https://doi.org/10.1177/1741143218753192>

- Cahill, S. M., Egan, B. E., Wallingford, M., Huber-Lee, C., & Dess-McGuire, M. (2015). Results of a school-based evidence-based practice initiative. *American Journal of Occupational Therapy*, 69, 1–5. <https://doi.org/10.5014/ajot.2015.014597>
- Cain, T. (2015). Teachers' engagement with research texts: Beyond instrumental, conceptual or strategic use. *Journal of Education for Teaching*, 41(5), 478–492. <https://doi.org/10.1080/02607476.2015.1105536>
- Cain, T. (2017). Denial, opposition, rejection or dissent: Why do teachers contest research evidence? *Research Papers in Education*, 32(5), 611–625. <https://doi.org/10.1080/02671522.2016.1225807>
- Cain, T. (2019). *Becoming a research-informed school: Why? What? How?* Routledge.
- Cirkony, C., Rickinson, M., Walsh, L., Gleeson, J., Salisbury, M., Cutler, B., Berry, M., & Smith, K. (2021). Beyond effective approaches: A rapid review response to designing professional learning. *Professional Development in Education*. Advance online publication. <https://doi.org/10.1080/19415257.2021.1973075>
- Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., Stoll, L., Willis, B., & Burns, H. (2017). *Evidence-informed teaching: An evaluation of progress in England*. Department of Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/625007/Evidence-informed_teaching_-_an_evaluation_of_progress_in_England.pdf
- Cooper, A., & Levin, B. (2013). Research use by leaders in Canadian school districts. *International Journal of Education Policy & Leadership* 8(7). <https://files.eric.ed.gov/fulltext/EJ1016293.pdf>
- Cordingley, P. (2015). The contribution of research to teachers' professional learning and development. *Oxford Review of Education*, 41(2), 234–252. <https://doi.org/10.1080/03054985.2015.1020105>
- Creed, P. A., Hood, M., Praskova, A., & Makransky, G. (2016). The career distress scale: Using Rasch measurement theory to evaluate a brief measure of career distress. *Journal of Career Assessment*, 24(4), 732–746. <https://doi.org/10.1177/1069072715616126>
- Curtis, D. D. (2005). Comparing classical and contemporary analyses and Rasch measurement. In S. Alagumalai, D. D. Curtis, & N. Hungi (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 179–195). Springer.
- Dagenais, C., Lysenko, L., Abrami, P. C., Bernard, R. M., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of empirical research. *Evidence & Policy*, 8(3), 285–309. <https://doi.org/10.1332/174426412X654031>
- Evans, C., Waring, M., & Christodoulou, A. (2017). Building teachers' research literacy: Integrating practice and research. *Research Papers in Education*, 32(4), 403–423. <https://doi.org/10.1080/02671522.2017.1322357>
- Figgis, J., Zubrick, A., Butorac, A. & Alderson, A. (2000). Backtracking practice and policies to research. In Department of Education, Training and Youth Affairs (Ed.), *The impact of educational research* (pp. 279–374). https://genbase.iiep.unesco.org/workspace/applis/epidoc/fichiers/EPIDOC/20264_G020264.pdf
- Finnigan, K. S., & Daly, A. J. (Eds.). (2014). *Using research evidence in education: From the school-house door to Capitol Hill*. Springer.
- Gitomer, D. H., & Crouse, K. (2019). *Studying the use of research evidence: A review of methods*. William T Grant Foundation.
- Godfrey, D., & Brown, C. (Eds.). (2019). *An ecosystem for research-engaged schools: Reforming education through research*. Routledge.
- Godfrey, D. (2016). Leadership of schools as research-led organizations in the English educational environment: Cultivating a research engaged school culture. *Educational Management Administration & Leadership*, 44(2), 301–321. <https://doi.org/10.1177/1741143213508294>
- Gorard, S. (Ed.). (2020). *Getting evidence into education*. Routledge.
- Hayes, A. F., & Coultts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability. *But. Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Hill, J. (2022). Who is facilitating research use in education systems? In OECD (ed.), *Who cares about using education research in policy and practice? Strengthening research engagement, educational research and innovation* (pp. 74–99). OECD Publishing. <https://doi.org/10.1787/d7ff793d-en>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60. <http://www.ejbrm.com/vol6/v6-i1/v6-i1-papers.htm>

- Joram, E., Gabriele, A. J., & Walton, K. (2020). What influences teachers' "buy-in" of research? Teachers' beliefs about the applicability of educational research to their practice. *Teaching and Teacher Education*, 88, 102980. <https://doi.org/10.1016/j.tate.2019.102980>
- Judkins, M., Stacey, O., McCrone, T., & Inniss, M. (2014). *Teachers' use of research evidence: A case study of United Learning schools*. National Foundation for Educational Research. <https://www.nfer.ac.uk/publications/IMUL01/IMUL01.pdf>
- Lawlor, J., Mills, K., Neal, Z., Neal, J. W., Wilson, C., & McAlindon, K. (2019). Approaches to measuring use of research evidence in K-12 settings: A systematic review. *Educational Research Review*, 27, 218–228. <https://doi.org/10.1016/j.edurev.2019.04.002>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Lysenko, L., Abrami, P., Bernard, R., Dagenais, C., & Janosz, M. (2014). Educational research in educational practice: Predictors of use. *Canadian Journal of Education*, 37(2), 1–26. <http://www.jstor.org/stable/canajeducrevucan.37.2.06>
- Malin, J. R., Brown, C., Ion, G., van Ackeren, I., Bremm, N., Luzmore, R., Flood, J., & Rind, G. M. (2020). World-wide barriers and enablers to achieving evidence-informed practice in education: What can be learnt from Spain, England, the United States, and Germany? *Humanities and Social Sciences Communications*, 7(1), 1–14. <https://doi.org/10.1057/s41599-020-00587-8>
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(3), 1–16. <https://pubmed.ncbi.nlm.nih.gov/18753691/>
- Martinovic, D., Wiebe, N., Ratkovic, S., Willard-Holt, C., Spencer, T., & Cantalini-Williams, M. (2012). 'Doing research was inspiring': Building a research community with teachers. *Educational Action Research*, 20(3), 385–406. <https://doi.org/10.1080/09650792.2012.697402>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42), 1–12. <https://doi.org/10.1186/1748-5908-6-42>
- Mills, M., Mockler, N., Stacey, M., & Taylor, B. (2021). Teachers' orientations to educational research and data in England and Australia: Implications for teacher professionalism. *Teaching Education*, 32(1), 77–98. <https://doi.org/10.1080/10476210.2020.1843617>
- Mincu, M. (2014). Inquiry paper 6: Teacher quality and school improvement – What is the role of research? In the BERA/RSA (Eds.), *The role of research in teacher education: Reviewing the evidence*. BERA. <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-RSA-Interim-Report.pdf>
- Neal, Z., Lawlor, J., Neal, J. W., Mills, K., & McAlindon, K. (2019). Just Google it: Measuring schools' use of research evidence with internet search results. *Evidence & Policy*, 15(1), 103–123. <https://doi.org/10.1332/174426418X15172392413087>
- Nelson, J., & Campbell, C. (2019). Using evidence in education. In A. Boaz, H. Davies, A. Fraser, & S. Nutley (Eds.), *What works now? Evidence-informed policy and practice revisited* (pp. 131–149). Policy Press.
- Nelson, J., Mehta, P., Sharples, J., & Davey, C. (2017). *Measuring teachers' research engagement: Findings from a pilot study*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Research_Use/NFER_Research_Use_pilot_report_-_March_2017_for_publication.pdf
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Organisation for Economic Co-operation and Development [OECD] (2022). *Who cares about using education research in policy and practice? Strengthening research engagement, educational research and innovation*. OECD Publishing. <https://doi.org/10.1787/d7ff793d-en>
- Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46(1), 1–18. <https://doi.org/10.1348/014466506X96931>
- Parker, B., Steele, T., Rose, V., & Taylor, D. (2020). *Getting evidence moving in schools (GEMS): Investigation paper*. Evidence for Learning. <https://evidenceforlearning.org.au/research-and-evaluation/investigations-and-insights/getting-evidence-moving-in-schools-gems/>

- Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C. C., & Allen, A.-R. (2017). How school and district leaders access, perceive, and use research. *AERA Open*, 3(2), 1–17. <https://doi.org/10.1177/2332858417705370>
- Penuel, W. R., Briggs, D.C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C. C., & Allen, A.-R. (2016). *Findings from a national survey of research use among school and district leaders*. National Center for Research in Policy and Practice. <https://files.eric.ed.gov/fulltext/ED599966.pdf>
- Plant, B., Boulet, M., & Smith, L. (2022). *A behavioural approach to understanding and encouraging quality use of research evidence in Australian schools: Final report*. BehaviourWorks Australia, Monash University. <https://doi.org/10.26180/21530658.v1>
- Poet, H., Mehta, P., & Nelson, J. (2015). *Research use in schools: Survey, analysis and guidance for evaluators*. National Center for Research in Policy and Practice.
- Prendergast, S., & Rickinson, M. (2019). Understanding school engagement in and with research. *Australian Educational Researcher*, 46, 17–39. <https://doi.org/10.1007/s13384-018-0292-9>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.
- Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16(2), 234–243. <https://doi.org/10.1111/j.1365-2702.2006.01573.x>
- Rickinson, M., Cirkony, C., Walsh, L., Gleeson, J., & Salisbury, M. (2021). *Towards quality use of research evidence in education: Discussion paper*. Monash University. <https://doi.org/10.26180/14071571.v2>
- Rickinson, M., Gleeson, J., Walsh, L., Cutler, B., Cirkony, C., & Salisbury, M. (2021). *Research and evidence use in Australian schools: Survey, analysis and key findings*. Monash University. <https://doi.org/10.26180/14445663>
- Rickinson, M., Gleeson, J., Walsh, L., Salisbury, M., Cutler, B., & Cirkony, C. (2021). *Using research well in Australian schools: Discussion paper*. Monash University. <https://doi.org/10.26180/14783637.v2>
- Rickinson, M., Walsh, L., Gleeson, J., Cutler, B., Cirkony, C., & Salisbury, M. (2022). Using research well in educational practice. In OECD (ed.), *Who cares about using education research in policy and practice? Strengthening research engagement, educational research and innovation* (pp. 182–199). OECD Publishing. <https://doi.org/10.1787/65aac033-en>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sachs, J. (2016). Teacher professionalism: Why are we still talking about it? *Teachers and Teaching*, 22(4), 413–425. <https://doi.org/10.1080/13540602.2015.1082732>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modelling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–337. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates.
- Smith, E. V. (2002). Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205–231. <https://pubmed.ncbi.nlm.nih.gov/12011501/>
- Stoll, L., Greany, T., Coldwell, M., Higgins, S., Brown, C., Maxwell, B., Stiel, B., Willis, B., & Burns, H. (2018). *Evidence-informed teaching: Self-assessment tool for teachers*. Chartered College of Teaching. <https://discovery.ucl.ac.uk/id/eprint/10042947>
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*, 7(S1), 22–26. <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>

- Thomson, S., & Hillman, K. (2019). *The teaching and learning international survey 2018. Australian report volume 1: Teachers and school leaders as lifelong learners*. Australian Council for Educational Research. <https://research.acer.edu.au/talis/6/>
- Tripney, J., Kenny, C., & Gough, D. (2014). Enabling the use of research evidence within educational policymaking in Europe: Lessons from the EIPEE Project. *European Education*, 46(1), 55–74. <https://doi.org/10.2753/EUE1056-4934460103>
- van Schaik, P., Volman, M., Admiraal, W., & Schenke, W. (2018). Barriers and conditions for teachers' utilisation of academic knowledge. *International Journal of Educational Research*, 90, 50–63. <https://doi.org/10.1016/j.ijer.2018.05.003>
- Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, 36(2), 299–316. <https://doi.org/10.1080/01411920902919257>
- Walker, M., Nelson, J., Bradshaw, S., & Brown, C. (2019). *Teachers' engagement with research: What do we know?* NFER. <https://www.nfer.ac.uk/teachers-engagement-with-research-what-do-we-know-a-research-briefing/>
- Wetzel, E., & Carstensen, C. H. (2014). Reversed thresholds in partial credit models: A reason for collapsing categories? *Assessment*, 21(6), 765–774. <https://doi.org/10.1177/1073191114530775>
- White, S., Down, B., Mills, M., Shore, S., & Woods, A. (2021). Strengthening a research-rich teaching profession: An Australian Study. *Teaching Education*, 32(3), 338–352. <https://doi.org/10.1080/10476210.2020.1737666>
- White, S., Nuttall, J., Down, B., Shore, S., Woods, A., Mills, M., & Bussey, K. (2018). *Strengthening a research-rich teaching profession for Australia*. Australian Teacher Education Association (ATEA), Australian Association for Research in Education (AARE), & Australian Council of Deans of Education, (ACDE). <https://www.aare.edu.au/assets/documents/Strengthening-a-research-rich-teaching-profession-FOR-RESEARCH-PAGE-v2.pdf>
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9(1), 1–26. <https://doi.org/10.1177/014662168500900101>
- Wieser, C. (2018). Evidence and its integration into teacher knowledge: Foucaultian perspectives to link research knowledge and teaching. *Journal of Education for Teaching*, 44(5), 637–650. <https://doi.org/10.1080/02607476.2018.1516352>
- Williams, D., & Coles, L. (2007). Teachers' approaches to finding and using research evidence: An information literacy perspective. *Educational Research*, 49(2), 185–206. <https://doi.org/10.1080/00131880701369719>
- Winch, C. (2017). *Teachers' know-how: A philosophical investigation*. Wiley.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Joanne Gleeson¹  · Blake Cutler¹  · Mark Rickinson¹  · Lucas Walsh¹  ·
John Ehrich²  · Connie Cirkony³  · Mandy Salisbury¹ 

Blake Cutler
blake.cutler@monash.edu

Mark Rickinson
mark.rickinson@monash.edu

Lucas Walsh
lucas.walsh@monash.edu

John Ehrich
john.ehrich@mq.edu.au

Connie Cirkony
connie.cirkony@utas.edu.au

Mandy Salisbury
mandy.salisbury1@monash.edu

¹ Faculty of Education, Monash University, 19 Ancora Imparo Way, Clayton, VIC 3800, Australia

² School of Education, Macquarie University, 14 First Walk, Macquarie Park, Sydney, NSW 2109, Australia

³ Faculty of Education, University of Tasmania, Churchill Avenue, Hobart, TAS 7005, Australia