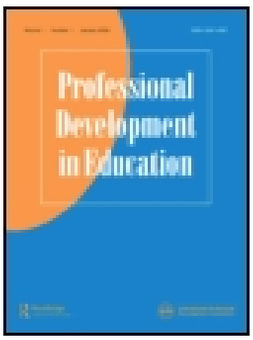

Professional Development in Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjie20

Development and validation of a survey on outcomes of professional learning

Chase R. Nordengren

To cite this article: Chase R. Nordengren (2020): Development and validation of a survey on outcomes of professional learning, Professional Development in Education

To link to this article: https://doi.org/10.1080/19415257.2020.1862276

Published online: 16 Dec 2020.

PROFESSIONAL DEVELOPMENT IN EDUCATION
https://doi.org/10.1080/19415257.2020.1862276

Routledge
Taylor & Francis Group

ARTICLE

Development and validation of a survey on outcomes of professional learning

Chase R. Nordengren

Professional Learning, NWEA, Portland, OR, USA

ABSTRACT

Professional learning evaluators struggle to balance the pursuit of rigorous outcomes with the time and money that some effective measures entail. While many evaluations rely on surveys for their ease of administration and application, little research has explored how surveys can measure multiple constructs beyond participant satisfaction as part of an evaluation framework. This study validates a survey of professional learning outcomes based on Guskey's framework of professional development evaluation. The 23-item survey was administered to 1,351 teachers in one district in the United States who receive professional learning interventions related to assessment use and assessment literacy. Using exploratory and confirmatory factor analyses, the study demonstrates the discriminatory validity of four constructs: knowledge gained from professional learning, skills used in the classroom, attitudes about professional learning, and beliefs about assessment. The study provides additional credence to the relevance of Guskey's framework for evaluating professional learning interventions and insights regarding the utility of survey research in program evaluation activities.

ARTICLE HISTORY
Received 24 February 2020
Accepted 2 December 2020

KEYWORDS
Professional learning; survey research; program evaluation; factor analysis

Introduction

Evaluation studies attempting to link professional learning to concrete outcomes for teachers and students face significant methodological and logistic challenges. While efforts in recent years have linked some program elements to changes in student outcomes (Darling-Hammond *et al.* 2017), research is generally unable to inform professional learning designers on the best ways to structure and deliver a particular set of content. Additionally, because the unique impact of professional learning on student outcomes is difficult to isolate (Guskey 2002a, Holloway 2006, Hirsh 2013), program evaluators often turn to proxy measures that link professional learning to its more direct outcomes on those who participate.

In the United States, professional learning is often a unique mix of content and methods in each district and school where it takes place; therefore, professional learning evaluation necessitates concrete evaluation tools that produce context-specific evidence with high validity and reliability. This study seeks to validate one such measure: a survey designed to gather participants' self-perceptions of the knowledge gained, skills used, and attitudes and beliefs changed as the result of participating in professional learning. Based on a survey administered following a program of professional learning about assessment use, this measurement study illustrates how general concepts of effectiveness in professional learning can be applied to specific knowledge, skills, and attitudes that a professional learning program seeks to affect.

This study asks:

CONTACT Chase R. Nordengren chase.nordengren@nwea.org Professional Learning, NWEA, Portland, OR, USA

(1) What is the underlying structure of a survey on professional learning effectiveness measuring factors related to Guskey's (2000) framework?
(2) Do each of the factors identified by the prior analysis demonstrate adequate discriminant validity?

Theoretical framework

Though professional learning is a critical component of educator quality and many school improvement initiatives, measuring its impact is far from a straightforward exercise. Because the impact of professional learning reaches students only indirectly through their educators, the likelihood of finding statistically significant impacts on student learning in the short term is slim. As is true in many other realms of educational best practice (Lortie-Forgues and Inglis 2019), hardly any experimental or quasi-experimental professional learning studies find significant impacts (Yoon *et al.* 2007, Gersten *et al.* 2014).

Alternative measures of effectiveness provide the opportunity to save cost, increase the likelihood of finding actionable information, and serve as both a more accurate form of summative evaluation and a source of formative insights that can support program improvement. Among these studies, surveys are a primary methodological tool: they are typically inexpensive to administer, allow inquiry into many types of impacts, and place minimal burden on participants. However, specific event surveys focussed only on participant satisfaction redirect focus from a systematic focus on programs and processes necessary for a fully formed evaluation effort (Killion 2018).

Survey methods themselves are also regularly criticised. Primarily, methodologists warn of the risks of respondents self-reporting data (Tourangeau *et al.* 2000): respondents may mis-understand questions, incorrectly estimate or guess the time span between key events, forget to report important details, or apply different levels of sensitivity on items asking them to assess relative benefits or downsides. Additional concerns around survey methods include variations in response due to a lack of trust from participants (Jimerson 2016), the possibility of a 'social desirability bias' in survey responses (Kreuter *et al.* 2008), and the perceived low quality of many survey items (Desimone and Le Floch 2004).

Among common evaluation methods, Desimone (2009) argues, surveys elicit the most criticism. However, Desimone questions whether 'everyone in the field has similar notions of what those limitations are and how they compare with the strengths and limitations of' other approaches (Desimone 2009, p. 188). Further, Desimone argues, many critiques of the accuracy of self-report data are based on methodologically weak studies completed in the 1960s and 1970s, using observations designed to identify different types of behaviours than the surveys with which they were matched (Desimone 2009).

Advancements in survey methodology over the last 30 years, spurned by the growth of online survey platforms and development of research areas like cognitive aspects of survey measurement (CASM), have driven additional rigour to survey work (Newcomer and Triplett 2015, Sha and Gabel 2020). Survey methodologists are deeply engaged in reducing the risks identified by critics through effective design of survey items (e.g., Fowler 1995, Presser *et al.* 2004, Saris and Gallhofer 2014), cognitive interviews with participants similar to future respondents (Miller *et al.* 2014), understanding the tradeoffs associated with various forms of survey implementation (e.g., Baker *et al.* 2004, Kreuter *et al.* 2008, Newcomer and Triplett 2015), modelling measurement error quantitatively to reject poorly-framed items (Biemer 2004, 2017) and other techniques.

As in many fields of applied research, using a variety of measures of effectiveness and matching those methods to the research questions being asked results in the best-informed judgements. Given that policymakers and planners sometimes act before all evidence is in, however, each instrument in an evaluation strategy must have its own robust understanding of the constructs it measures, its strengths in speaking to overall effectiveness, and its weaknesses relative to other methods. To do so in the context of professional learning requires a full understanding of the space: the ways that

teachers can both directly influence students as a result of their professional learning and indirectly drive school cultures and common practices to better support the learning of all students.

In summarising several models for effective evaluations, Guskey (2000) provides a five-level schema for collecting and analysing data on the effectiveness of professional learning. This model is designed to capture the influences that professional development has on practitioners in a system that ultimately leads to more effective instruction, strengthened school cultures, and better outcomes for students. Guskey's five levels include the following:

(1) *Participants' reactions* (whether participants enjoyed the learning, felt it was time well spent, and found the information useful and helpful).
(2) *Participants' learning* (whether participants acquired the intended knowledge and skills from the professional learning).
(3) *Organisational support and change* (how professional learning impacted the climate, procedures, resources, and problems of the school and district).
(4) *Participants' use of new knowledge and skills* (whether and how participants applied the knowledge and skills learned in classroom and building practice).
(5) *Student learning outcomes* (how professional learning impacted students both cognitively and affectively, and in psychomotor skills and behaviours).

As part of a complete evaluation strategy that gathers evidence on all five levels, tools that can measure participants' learning and use of new skills serve to help connect the immediate impacts of a professional learning activity to long-term impacts on student learning outcomes.

While Guskey's framework provides a complete picture of professional learning impacts and aligns with previous work on the effect of other types of training programs (Kirkpatrick 1977), levels of the framework often require the use of multiple instruments targeting a variety of stakeholders. Gathering information on student outcomes, for example, requires the collection and analysis of student records (Guskey 2000). Similarly, gathering information on organisational support and change requires an extensive understanding of culture and context usually attained through interviews, focus groups, or other more intensive methods (Guskey 2000). While these approaches are highly meritorious, and important elements of any balanced professional learning plan, they are also by their nature resource intensive. These demands accrue not only for evaluators but also for educators, administrators, and schools.

Surveys offer particular utility when understanding participant attitudes and beliefs in relationship to changes in teaching practice. Whitley *et al.* (2019) apply this approach, using a survey instrument to demonstrate the alignment between teacher beliefs around differentiated instruction and self-reports of teacher practices in differentiated instruction. Brookhart *et al.* (2009) use a survey of attitudes and beliefs alongside student achievement measures to understand the impact of professional development on formative assessment practices. Buzick *et al.* (2019) use a series of surveys to understand the relationship between knowledge, attitudes, and data use practices on Common Core-aligned assessments. While these efforts cannot understand the causal order of changes in beliefs versus changes in practices (and while Guskey (2002b) challenges the conventional wisdom that beliefs precede practices), the use of survey methods allows teaching practice to be understood in the context of what teachers believe and feel about what makes effective teaching.

While many evaluators have used Guskey's framework to understand evidence from professional learning surveys, relatively little published work has addressed how evaluators can match Guskey's levels to specific survey constructs and items. Using Guskey's framework as a holistic lens, this paper looks to provide evidence of the relationship between four survey constructs, derived from Guskey's framework, in the context of a specific organisational evaluation of professional learning. NWEA provides a variety of professional learning content on the use of its assessment products, assessment literacy, formative instructional practices, and other topics of relevance to K–12 teachers and administrators. Additionally, that professional learning is provided in a variety of forms,

including in-person workshops, virtual coaching sessions, and small-segment, blended learning experiences. As part of a professional learning evaluation strategy by NWEA, the survey examined here is designed to provide large-scale evidence of the overall impact of professional learning on the knowledge, skills, attitudes, and beliefs of educators in a variety of district contexts. While some instruments 'go deep' to explore opportunities for specific professional learning implementations to improve in their given contexts, this survey instead 'goes broad' to gather information from the most professional learning participants possible with limited impact on participants' time.

Methods

As part of a comprehensive evaluation strategy for a professional learning organisation, the survey discussed in this study is the second of four data collection elements, including needs assessment, classroom observations, and a participant portfolio. I developed each of these instruments in response to the organisation's theory of change Nordengren (2020), which outlines the impacts the organisation intends to have on teaching and learning and describes intermediate measures and outcomes that support inferences about achieving those outcomes. The instruments also account for the unique content of the organisation's professional learning: a focus on assessment use, assessment literacy, and formative assessment. The theory of change draws heavily from Guskey (2000, 2002a) to understand how to meaningfully capture the immediate and ultimate impacts from professional learning.

Each of the four data collection elements are designed to complement the data gathered from each other, providing opportunities for triangulation. However, the survey was also developed acknowledging the reality that some district and school participants may choose to only participate in the survey to minimise the time commitments of educators and administrators. This measurement study intends to account for the second possibility by exploring the extent to which the survey can independently represent constructs of meaning and interest to participating schools.

Any professional learning evaluation strategy should account for the content delivered to educators and the contexts in which the content was delivered. While some of the specific items outlined in this survey may have utility for other professional development, it is unlikely the survey will apply to another organisation's work wholescale. While this study may not provide a 'packaged' survey instrument for use elsewhere, it does provide useful constructs to include in such a survey, demonstrate a method for relating those constructs to Guskey's framework, and suggest the steps evaluators should take in demonstrating the discriminant validity of that survey.

Instrument

The Perceptions and Practices Survey is a 23-item measure consisting of four hypothesised constructs:

- The *attitudes* participants have towards professional learning and its value.
- Participants' *knowledge* in assessment use and assessment literacy.
- The *beliefs* participants have about assessments and assessment use.
- The *skills* participants regularly use in classroom instruction related to assessment use and assessment literacy.

These constructs derive from Guskey's (2000) framework and were adapted to the specific topics addressed in professional learning. In general, *knowledge* relates to participants' learning, *attitudes* and *beliefs* relate to participants' reactions to professional learning, and *skills* relates to participants' use of new knowledge and skills from professional learning on a regular basis (i.e., over the last month). Items on *attitudes* and *beliefs* may also provide insights into organisational support and change to the extent they represent the outcomes of appropriate or insufficient organisational

support for implementing a new initiative (in this case, a new assessment). By monitoring changes in these constructs over multiple administrations, the survey provides a measure of changes in knowledge, skills, attitudes and beliefs that correlate with professional learning participation.

Here, I separate *attitudes* and *beliefs* deliberately to better understand the relationship between two separate ideas: beliefs about assessment and attitudes towards this specific set of professional learning experiences. Unlike attitudes, participants' beliefs about assessment are pre-formed: they come from knowledge and experiences with all kinds of assessments throughout their personal and professional lives. Professional learning seeks to change these beliefs as part of an overall effort to change instructional practice. The attitudes described here are instead feelings about the professional learning they received during the year; they are an indicator of the potential willingness of participants to engage in more professional learning in the future. While addressed in similar ways by Guskey, this study looks to understand the extent to which these constructs can be measured separately from one another.

The survey as administered is provided in Appendix A.

Survey development

Based on this framework, I developed an initial survey with 28 items related to intended professional learning outcomes. In spring 2019, 12 teachers from a district in the northwestern United States who had received professional learning completed cognitive interviews (Miller *et al.* 2014) with the initial version of the survey. In these interviews, respondents were taken through the survey and asked to think aloud about the knowledge, memories, and ideas they were drawing upon to form answers.

Based on outcomes from cognitive interviews, I removed five items, reworded three items, and substantially revised the instructional text about items.

Survey participants and characteristics

Participants in the validation study were teachers across grade levels in an urban school district in the western United States. These teachers participated in NWEA professional learning during the 2019–20 school year and used the NWEA MAP Growth assessment. The survey was distributed by the district electronically in October 2019 to all teachers in ninety participating schools, and 1,351 complete responses were received.

The survey was provided electronically through SurveyGizmo. Use of web-based surveys may reduce the risk of social desirability bias compared to other methods (Kreuter *et al.* 2008). Individual participant identifiers were not collected to ensure participant anonymity and high participation. One disadvantage of this approach is an inability to prevent participants from submitting duplicate responses. Additionally, because many teachers completed the survey from the same school buildings, other approaches for protecting against duplicate responses (such as IP address monitoring) are also unworkable. The sponsoring organisation and district institutional review boards reviewed and provided oversight on confidentiality procedures.

Demographics for participants are reported in Table 1 and are generally in line with demographics reported by the district for teachers overall. Question sets were randomised to avoid order effects. The survey system did not allow for the recording of partial responses.

Analysis

This study sought both to conduct exploratory analyses that could understand the factor structure of the items in the survey and to validate the results of that analysis. To do so, I conducted an exploratory factor analysis and confirmatory factor analysis using separate analytic samples. Exploratory factor analysis (EFA) is a technique for discovering and defining the latent constructs

Table 1. Demographics of survey participants.

		n	%
Gender			
	Female	1028	76.1
	Male	197	14.6
	Other/Not specified	126	9.3
Race			
	Asian	82	6.1
	Native Hawaiian or Other Pacific Islander	17	1.3
	Black/African American	82	6.1
	White	781	57.8
	Hispanic/Latino	119	8.8
	American Indian/Alaskan Native	6	0.4
	Not specified	264	19.5
Years in Education			
	< 3	179	13.2
	3–5	198	14.7
	6–10	235	17.4
	11–15	233	17.2
	> 15	506	37.5

that underlie a set of variables, such as survey responses, to provide the preliminary basis for a causal analysis between and among those latent constructs (Loehin and Beaujean 2017). Confirmatory factor analyses (CFA) then seek to validate the factor structure hypothesised by the EFA analysis.

I divided the total data set in half by order of receipt, with even-numbered responses used for exploratory factor analyses and odd-numbered responses used for confirmatory factor analyses. This provided a final sample size of 675 responses for EFA analyses and 676 responses for CFA analyses.

Results

Exploratory factor analysis [EFA]

As recommended by Loehin and Beaujean (2017), I compared two methods for identifying the number of factors appropriate to this data set: the Kaiser-Guttman rule and parallel analysis. In both methods, potential numbers of factors are compared with one another using eigenvalues, a statistic representing the amount of variance accounted for in the overall model by the addition of a new factor. Plotting these on a graph known as a scree plot represents the iterative value of adding each additional factor to the model. The Kaiser-Guttman rule suggests, by past precedent, that the number of factor models with eigenvalues greater than 1.0 represents the number of nontrivial factors present in the data. Parallel analysis instead develops random data sets with the same number of items as the tested data set, estimates eigenvalues for these data sets, and compares their values to the tested data set. Factors that outperform these 'random' data sets are retained.

Eigenvalues for this data set and data simulated using parallel analysis are shown in Table 2 and presented as a scree plot in Figure 1. On the plot, the solid horizontal line represents the Kaiser-Guttman rule cut-off (eigenvalues greater than 1), indicating three factors are appropriate. The results of the parallel analysis are understood as the number of factors in which observed data shows a higher eigenvalue than simulated data, indicating four factors are appropriate.

Given this finding, I estimated two separate EFAs specifying a four-factor model and a three-factor model using R's *lavaan* package (a set of software tools used to model the data), following procedures in Loehin and Beaujean (2017). In each analysis, oblique rotation and minimum residual factoring were used; these choices conservatively assume non-normal distribution of

Table 2. Eigenvalues for first 10 factors in real and simulated models.

Factor	Real Data	Simulated Data
1	11.60	0.39
2	1.93	0.31
3	1.51	0.26
4	0.82	0.23
5	0.18	0.19
6	−0.03	0.16
7	−0.08	0.13
8	−0.15	0.10
9	−0.16	0.08
10	−0.18	0.05

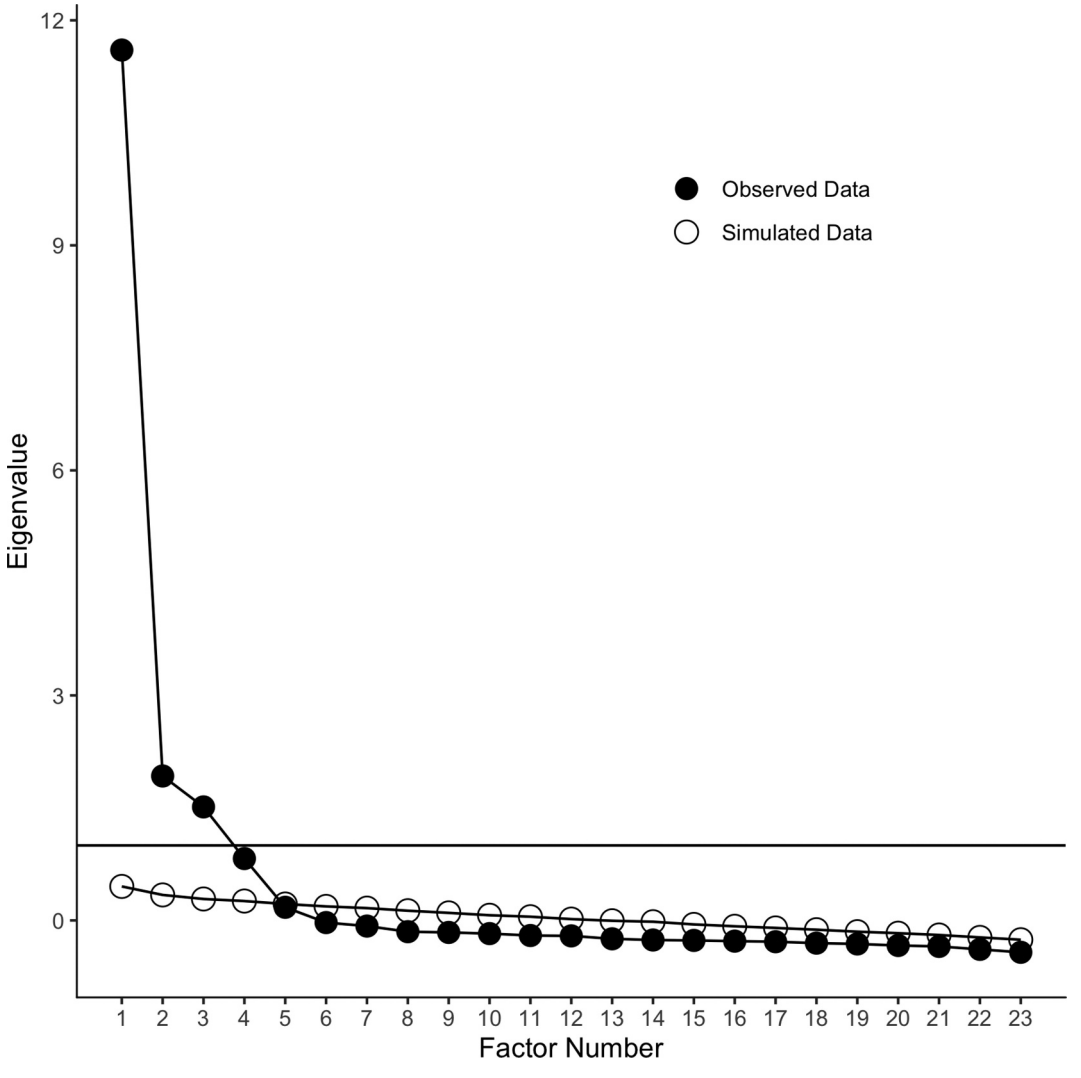

Figure 1. Scree plot results of parallel analysis.

responses and correlated factors. I selected a minimum factor loading threshold of.40 (Netemeyer *et al.* 2003).

Table 3. Four-factor and three-factor model of survey items.

	3-Factor Solution			4-Factor Solution			
Item	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
1	**0.795**	0.049	0.073	**0.796**	0.052	0.059	0.020
2	**0.827**	0.024	0.009	**0.823**	0.027	−0.001	0.016
3	**0.816**	0.017	0.055	**0.814**	0.021	0.034	0.028
4	**0.897**	0.001	0.045	**0.896**	0.005	0.041	0.009
5	**0.899**	0.028	−0.029	**0.894**	0.030	−0.034	0.009
6	**0.930**	−0.007	−0.041	**0.924**	−0.006	−0.049	0.013
7	**0.824**	−0.011	0.017	**0.825**	−0.008	0.051	−0.039
8	−0.070	**0.867**	0.042	−0.064	**0.863**	0.061	−0.013
9	−0.019	**0.889**	0.039	−0.016	**0.883**	0.010	0.043
10	0.083	**0.881**	−0.056	0.080	**0.874**	−0.070	0.026
11	0.004	**0.918**	0.014	0.008	**0.911**	0.010	0.014
12	−0.004	**0.908**	0.023	0.001	**0.903**	0.042	−0.014
13	0.059	**0.887**	−0.043	0.061	**0.881**	−0.008	−0.032
14	0.059	−0.032	**0.830**	0.096	−0.020	**0.735**	0.128
15	0.041	0.039	**0.778**	0.076	0.048	**0.758**	0.049
16	−0.044	0.014	**0.766**	−0.020	0.013	**0.890**	−0.078
17	0.045	0.001	**0.831**	0.080	0.007	**0.822**	0.046
18	−0.137	0.026	**0.785**	−0.113	0.029	**0.842**	−0.013
19	0.070	0.081	**0.602**	0.046	0.071	−0.020	**0.790**
20	−0.048	0.094	**0.627**	−0.087	0.084	−0.020	**0.836**
21	0.091	−0.065	**0.710**	0.071	−0.090	0.060	**0.843**
22	0.179	−0.026	**0.617**	0.183	−0.026	0.156	**0.561**
23	−0.251	0.208	0.347	−0.259	0.210	0.036	0.380

Bold values are above the decision-making threshold for inclusion in a given factor (0.4).

The results of the three-factor model (Table 3) showed that 22 of 23 items uniquely loaded on individual factors. When examining the factors, items written to describe teacher knowledge are contained within in Factor 1, items describing teacher use of skills are contained within Factor 2, with Factor 3 containing both items describing attitudes about professional learning and beliefs about assessment. Item 23 did not load onto any of the factors over the minimum loading threshold of .40. Correlations between factors (Table 4) were between 0.57 and 0.62. The three-factor model explains 65.7% of the cumulative variance of the data.

The results of the four-factor model (Table 4) also showed that the same 22 items uniquely loaded on the individual factors. This structure looks substantially similar to the three-factor model, while separating items on attitudes about professional learning and items on beliefs about assessment into Factors 3 and 4, respectively. Correlations between factors (Table 4) were between 0.52 and 0.60. The four-factor model explains 69.3% of the cumulative variance of the data.

These results suggest the four-factor model provides modestly more value than the three-factor model in understanding the variance present in the data set. The four-factor model explains a larger percentage of the variance and shows reduced overall correlation between factors. Both models, however, show reasonable explanatory power and correlations that indicate the expected interrelationship between factors without major overlap. However, the relatively close nature of the two

Table 4. Correlations between factors in four-factor and three-factor models.

	3-Factor Solution			4-Factor Solution			
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	–			–			
Factor 2	.626	–		.569	–		
Factor 3	.569	.607	–	.589	.561	–	
Factor 4				.527	.522	.598	–

models suggests the need for confirmatory factor analysis to test the relative fit of each model to better understand each model's power to explain the data.

Confirmatory factor analysis [CFA]

The second phase of the analysis further examined the dimensionality of the survey using confirmatory factor analysis to compare a three- and four-factor model with a separate subsample of survey respondents. Given the low factor loadings for Item 23 in the EFA models, I excluded this item from further analysis. The revised four-factor structure is shown in Table 5. The three-factor structure differs from the four-factor structure by combining factors 3 and 4. Model fit was computed using maximum likelihood [MLR] estimation.

Model fit results from both models are presented in Table 6. The chi-square results for both the three-factor (χ^2 (206) = 1622.805, p < .001) and the four-factor (χ^2 (203) = 833.969, p < .001) models were significant, a result commonly associated with lack of model fit. However, chi-square values are prone to over-sensitivity in larger sample sizes. To adjust for this over-sensitivity, I used χ^2/df ratios for both models. I also used three measures of model fit recommended in Loehin and

Table 5. Final factor structure.

Item Number	Item Description
Factor 1	*Skills*
1	Understand the skills students are ready to learn
2	Place students into instructional groups
3	Understand students' performance relative to each other
4	Make informed decisions about instructional next steps
5	Set goals for individual students
6	Set goals for your class
7	Set goals with your teaching team
Factor 2	*Knowledge*
8	Review and understand information about an individual student's performance on MAP Growth assessments
9	Relate results on MAP Growth to what a student knows and can do
10	Use MAP Growth assessment data to place students in instructional groups
11	Use MAP Growth assessment data to set actionable goals for individual students
12	Use MAP Growth assessment data to understand what a particular student is ready to learn next
13	Use MAP Growth assessment data to understand where you should focus whole-group instruction
Factor 3	*Attitudes about Professional Learning*
14	NWEA Professional Learning has helped me use MAP Growth data more in my practice.
15	NWEA Professional Learning has helped me better explain MAP Growth results to parents.
16	NWEA Professional Learning has helped me better understand how to adjust my instruction or professional practice for the culture, climate, and sensitivities of my district, school, classroom, and students.
17	NWEA Professional Learning has helped me better use formative assessment techniques to supplement data from MAP Growth.
18	NWEA Professional Learning has helped me better reflect on my professional practice.
Factor 4	*Beliefs About Assessment*
19	MAP Growth provides information I can use in the classroom.
20	MAP Growth shows how students have grown across academic years.
21	MAP Growth is worth the time our school spends administering it each year.
22	I have personally witnessed how MAP Growth can help students grow.

In the three-factor model, factors 3 and 4 are combined.

Table 6. Fit indices for three- and four-factor models.

Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA	$\Delta\chi^2$
Three-factor	1622.805***	206	7.875	.907	.896	.101	
Four-factor	833.969***	203	4.11	.957	.953	.068	788.84***

CFI: comparative fit index, TLI: Tucker-Lewis Index, RMSEA: Root Mean Square Error of Approximation. Significance level *p <.05, **p <.01, and ***p <.001.

Beaujean (2017): Root Mean Square Error of Approximation (RMSEA), comparative fit index (CFI), and the Tucker-Lewis Index (TLI).

In this analysis, the four-factor model showed better model fit according to general guidelines (Browne and Cudeck, 1992; Cangur and Ercan, 2015). Evaluating both models using RMSEA (value below 0.08 representing reasonable fit), CFI (values about 0.95 represent acceptable fit), and TLI (values above 0.95 represent acceptable fit), the four-factor model exceeded these thresholds, while the three-factor model did not. These values are also presented in Table 5. Finally, I used a chi-square difference test to compare the fit of the two models. This test was significant ($\Delta\chi^2$ (3) = 788.84, $p < .001$), suggesting the four-factor model provides a significant reduction in model misfit over the three-factor model.

These results confirm the hypothesised factor structure of the four separate factors as measures of independent states. While the factors relate to each other as expected, this relationship differs enough to suggest they measure separate constructs. Further, they support the hypothesis that a model that adds a fourth factor, which distinguishes attitudes about professional learning from attitudes about assessment more generally, is a better fit to the data.

Discussion

Significance

This study supports a line of work regarding the use of survey methods to understand professional learning effectiveness. While surveys cannot by themselves provide the depth necessary to understand interactions between professional learning content, participants, and the students they serve, they can provide the ability to examine certain causal links between processes and outcomes against large samples and across a variety of contexts (Desimone and Le Floch 2004). As part of the balanced evaluation strategy in which this survey is meant to play a part, the goal of this survey is to provide a high-level view of the impact of professional learning in a given setting, producing additional incentive for district stakeholders to pursue more in-depth evaluation work.

This study provides evidence supporting Guskey's (2000) framework and its efforts to disentangle the direct impacts of professional learning activities through, in this case, the four constructs of knowledge, skills, attitudes, and beliefs. The survey shows discriminant validity between these constructs, demonstrating the ability of a survey to measure four of Guskey's five elements of professional learning effectiveness in part. The study is the first step in a line of research exploring how the survey functions both independently and in tandem with other evaluation system elements.

The study demonstrates a localised and context-specific approach to understanding the impact of professional learning. Surveys used across generalised contexts have not provided concrete and definitive evidence around professional learning's impact (Guskey 2000). Despite the apparent 'consensus' around effective professional learning, built from several meta-analyses, differences between specific professional learning programs make it difficult to distinguish the elements that are causally essential from those that happen to correlate with essential elements (Sims and Fletcher-Wood 2020). The solution to this dilemma is ensuring professional learning evaluation is specific to the context of a particular professional learning implementation (McChesney and Aldridge 2019). This study demonstrates how general principles for measuring the impact of professional learning can be instrumentalized to meet the context of a specific professional learning organisation, its content, and its methods of delivery.

Limitations

This paper is limited by the lack of other outcomes to which survey data could be compared and evaluated for predictive validity. In particular, knowledge and the use of skills are constructs best measured through methods that are less reliant on self-report, such as assessments or observations

of classroom behaviour. My future work will leverage our organisation's work in these areas to correlate observation and portfolio work with survey results. However, practical realities in many environments – including the substantial time commitment entailed in more substantial evaluative tools and the perception among some educators of evaluation as a punitive enterprise – often limit the ability of evaluators to use other tools. Professional learning evaluators should continue to place a strong focus on practical and effective techniques for gathering data from participants that minimise the time asked from participants and that build collaborative and open relationships.

Needs of survey design also potentially limit the significance of this study's results. Because items within each hypothesised construct were grouped together using common stems rather than randomly distributed among a question set, some respondents might inadvertently cluster their responses more than they would under other circumstances. This design was necessary to ensure the readability of the survey and to fully differentiate between what participants know and the actions they take. Further, Schaeffer and Dykema (2004) suggest that careful control of question wording and ordering in designs like this one helps prevent respondents from confusing similar constructs. Nonetheless, future design considerations should seek to separate items in each construct from each other to the extent feasible and appropriate.

Conclusions

Program evaluations require a firm understanding of the tradeoffs between rigour and relevance. While survey methods rarely benefit from the strong causal conclusions and generalisability of controlled trials, they achieve advantages in expense and time that can be crucial to certain stakeholders, particularly those who wish to leverage an ongoing evaluation to inspire persistence among participants or otherwise motivate change. In the real-world settings in which most professional learning takes place, those responsible rarely ask for proof that something works; rather, they seek evidence of how the program may make meaningful impact on variables meaningful to them (Guskey 2002a).

Guskey's (2000) framework provides one prominent example of a set of relevant stakeholder variables on which program evaluations are based. To bring the framework into action, it must be accompanied by a set of feasible tools from which to gather that evidence. The survey described in this study successfully measures certain aspects of this framework from the perspective of professional learning participants. It functions both as a demonstration of the utility of the survey itself and the viability of survey methods to explore these questions more generally. Future work should continue applying Guskey's framework through a variety of tools, including online surveys, to best determine methods for gathering evaluation data without unduly burdening time-pressured teachers and administrators.

While much insight can come from survey tools, they cannot serve as the only word on the effectiveness of a professional learning program. As in any complex intervention, only a variety of evaluation measures – including both self-reported and externally observed outcomes linking to a concrete theory of action – can provide a complete picture of how professional learning impacts educators, systems, and students. Nevertheless, the process of constructing an evaluation strategy must begin from a source of immediate, actionable, and affordable insights. Effective surveys can serve to provide that immediate insight for educators and systems, potentially growing interest in and ownership of more complex evaluation strategies in the future.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Baker, R.P., Crawford, S., and Swinehart, J., 2004. Development and testing of web questionnaires. *In*: S. Presser, *et al*., eds. *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley, 361–384. doi:10.1002/0471654728.ch18.

Biemer, P., 2004. Modeling measurement error to identify flawed questions. *In*: S. Presser, *et al*., eds. *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley . 225–246. doi:10.1002/0471654728.ch12.

Biemer, P.P., 2017. *Total survey error in practice*. Hoboken, NJ: Wiley.

Brookhart, S.M., Moss, C.M., and Long, B.A., 2009. Promoting student ownership of learning through high-impact formative assessment practices. *Journal of multi disciplinary evaluation*, 6 (12), 42–67.

Browne, M.W. & Cudeck, R. 1992. "Alternative ways of assessing model fit", Sociological Methods & Research, 21, 230–258

Buzick, H.M., *et al*. 2019. *Teachers' views of their practices related to common core state standards-aligned assessments*. ETS Research Report Series. https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12277.

Cangur, S. & Ercan, I. 2015. "Comparison of model fit indices used in structural equation modeling under multivariate normality", Journal of Modern Applied Statistical Methods, 14, 152–167

Darling-Hammond, L., Hyler, M., and Gardner, M. 2017. *Effective teacher professional development*. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report.

Desimone, L.M., 2009. Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educational researcher*, 38 (3), 181–199. doi:10/c5738m

Desimone, L.M. and Le Floch, K.C., 2004. Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational evaluation and policy analysis*, 26 (1), 1–22. doi:10/bcgsws

Fowler, F.J., 1995. *Improving survey questions: design and evaluation*. Thousand Oaks, CA: SAGE Publications.

Gersten, R., *et al*., 2014. *Summary of research on the effectiveness of math professional development approaches (REL 2014–010)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory.

Guskey, T.R., 2000. *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.

Guskey, T.R., 2002a. Does it make a difference? Evaluating professional development. *Educational leadership*, 59 (6), 45–51.

Guskey, T.R., 2002b. Professional development and teacher change. *Teachers and teaching*, 8 (3), 381–391. doi:10/fnm4h7

Hirsh, S., 2013. The impact factor: why we can't neglect professional learning evaluation. *Journal of staff development*, 34 (5), 10–16.

Holloway, J.H., 2006. Connecting professional development to student learning gains. *Science educator*, 15 (1), 37-43.

Jimerson, J.B., 2016. How are we approaching data-informed practice? Development of the survey of data use and professional learning. *Educational assessment, evaluation and accountability*, 28 (1), 61–87. doi:10.1007/s11092-015-9222-9

Killion, J. 2018. *Assessing impact: evaluating professional learning*. Thousand Oaks, CA: Corwin Press.

Kirkpatrick, D.L., 1977. Evaluating training programs: Evidence vs. proof. *Training and development journal*, 31 (11), 9-12.

Kreuter, F., Presser, S., and Tourangeau, R., 2008. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public opinion quarterly*, 72 (5), 847–865. doi:10/b98xkm

Loehin, J.C. and Beaujean, A.A., 2017. *Latent variable models: an introduction to factor, path, and structural equation analysis*. 5th ed. New York, NY: Routledge. doi:10.4324/9781410609823.

Lortie-Forgues, H. and Inglis, M., 2019. Rigorous large-scale educational RCTs are often uninformative: should we be concerned? *Educational researcher*, 48 (3), 158–166. doi:10.3102/0013189x19832850.

McChesney, K. and Aldridge, J., 2019. Weaving an interpretivist stance throughout mixed methods research. *International journal of research & method in education*, 1–14. doi:10.1080/1743727x.2019.1590811.

Miller, K., *et al*., 2014. *Cognitive interviewing methodology*. Hoboken, NJ: Wiley. doi:10.4135/9781412983655.

Netemeyer, R.G., Bearden, W.O., and Sharma, S., 2003. *Scaling procedures: issues and applications*. Thousand Oaks, CA: SAGE.

Newcomer, K.E. and Triplett, T., 2015. Using surveys. *In*: K.E. Newcomer, H.P. Hatry, and J.S. Wholey, eds. *Handbook of practical program evaluation*. 4th ed. John Wiley & Sons, Inc, 344–382. doi:10.1002/9781119171386.ch14.

Nordengren, C. (2020), "Focusing squarely on students: A theory of change for NWEA professional learning", NWEA, https://www.nwea.org/resource-library/professional-learning/focusing-squarely-on-students-a-theory-of-change-for-nwea-professional-learning

Presser, S., *et al*., 2004. Methods for testing and evaluating survey questions. *In*: S. Presser, ed.. *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley, 1–22.

Saris, W.E. and Gallhofer, I.N., 2014. *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, NJ: Wiley.

Schaeffer, N.C. and Dykema, J., 2004. A multiple-method approach to improving the clarity of closely related concepts: distinguishing legal and physical custody of children. *In*: S. Presser, *et al*., eds. *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley, 475–502. doi:10.1002/0471654728.ch23.
Sha, M. and Gabel, T., 2020. *The essential role of language in survey research*. Research Triangle Park, NC : RTI Press.
Sims, S. and Fletcher-Wood, H., 2020. Identifying the characteristics of effective teacher professional development: A critical review. *School effectiveness and school improvement*, 1–17. doi:10.1080/09243453.2020.1772841.
Tourangeau, R., Rips, L.J., and Rasinski, K.A., 2000. *The psychology of survey response*. New York, NY: Cambridge University Press.
Whitley, J., *et al*., 2019. Implementing differentiated instruction: A mixed-methods exploration of teacher beliefs and practices. *Teachers and teaching*, 25 (8), 1043–1061. doi:10.1080/13540602.2019.1699782
Yoon, K.S., *et al*. 2007. *Reviewing the evidence on how teacher professional development affects student achievement*. https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf.

Appendix A. Perceptions and practices survey

In the past month, how often have you used data from MAP Growth assessments to (response options: none of the time/occasionally/sometimes/often):

(1) Understand the skills students are ready to learn
(2) Place students into instructional groups
(3) Understand students' performance relative to each other
(4) Make informed decisions about instructional next steps
(5) Set goals for individual students
(6) Set goals for your class
(7) Set goals with your teaching team

How confident are you that you can (response options: not at all confident/a little confident/somewhat confident/very confident):

(1) Review and understand information about an individual student's performance on MAP Growth assessments
(2) Relate results on MAP Growth to what a student knows and can do
(3) Use MAP Growth assessment data to place students in instructional groups
(4) Use MAP Growth assessment data to set actionable goals for individual students
(5) Use MAP Growth assessment data to understand what a particular student is ready to learn next
(6) Use MAP Growth assessment data to understand where you should focus whole-group instruction

To what extent do you agree or disagree with the following? (response options: strongly disagree/disagree/agree/strongly agree)
(1) NWEA Professional Learning has helped me use MAP Growth data more in my practice.
(2) NWEA Professional Learning has helped me better explain MAP Growth results to parents.
(3) NWEA Professional Learning has helped me better understand how to adjust my instruction or professional practice for the culture, climate, and sensitivities of my district, school, classroom, and students.
(4) NWEA Professional Learning has helped me better use formative assessment techniques to supplement data from MAP Growth.
(5) NWEA Professional Learning has helped me better reflect on my professional practice.

To what extent do you agree or disagree with the following?
(1) MAP Growth provides information I can use in the classroom.
(2) MAP Growth shows how students have grown across academic years.
(3) MAP Growth is worth the time our school spends administering it each year.
(4) I have personally witnessed how MAP Growth can help students grow.
(5) I can help every one of my students achieve academic growth.