Towards successful innovations in education: Development and validation of a multi-dimensional Innovative Work Behaviour Instrument

Peggy Lambriex-Schmitz¹ · Marcel R. Van der Klink^{1,2} · Simon Beusaert² ·
Monique Bijker¹ · Mien Segers²

Received: 27 February 2019 / Accepted: 16 January 2020 / Published online: 8 February 2020
© The Author(s) 2020

Abstract

Our study aims to develop and validate a multidimensional Innovative Work Behaviour instrument to measure teachers IWB. Current IWB conceptualizations and operationalizations need further attention. Existing measurements miss empirical evidence of the construct validity and moreover do not include a sustainability dimension. Based on a thorough and comprehensive conceptualization of IWB, we first adapted and extended the items of previously used instruments and developed items for a sustainability dimension. Second, we tested the construct validity of this newly developed multi-dimensional IWB instrument in a Dutch context, using Rasch techniques and Confirmatory Factor Analysis (CFA). The psychometric characteristics were examined in a sample of teachers working in vocational education (N = 440). The results revealed five dimensions of IWB, namely Opportunity Exploration, Idea Generation, Idea Promotion, Idea Realization (differentiated in two sub-dimensions: Criterion-based implementation and learning-based communication) and Idea sustainability (differentiated in two sub-dimensions: External dissemination and internal embedding). The dimensions were reliably measured (α .84–.94) using 44 items. This new instrument, which builds on recent conceptualizations of IWB as well as on the pivotal innovation models of West and Farr (1989) and Fullan (2007), offers for an international group of researchers a conceptually sound and valid tool to validate explanatory models of innovative teacher behaviour. For practitioners in the domain of educational innovation, it offers the opportunity to diagnose, in a sound way, the extent of the necessary conditions of teachers IWB for an educational innovation to succeed.

Keywords Innovative work behaviour · Successful innovations · Vocational teachers · Scale development · Validation · Sustainable innovations

✉ Peggy Lambriex-Schmitz
peggy.lambriex@zuyd.nl

Introduction

Higher and vocational education experience the sheer necessity to innovate. Rapidly changing labour markets require the acquisition of skills and knowledge that are often absent in education programs (World Economic Forum, 2016). Moreover, there is an ongoing debate on safeguarding the quality of education with regard to improving instructional methods, assessments, and their impact on teachers' competences (Organization for Economic Cooperation and Development, 2009; 2018 [OECD]). Today's innovations in education concern, for example, integrating twenty-first century skills or soft skills such as communication, collaboration and flexibility in curricula, to support students' entrepreneurial capacities (Schleicher, 2012; Sias et al., 2017). Also, new technologies in the classroom that foster new ways of teaching and learning are implemented (Bourgonjon et al., 2013; Eteokleous, 2008).

However, the need for innovations in education strongly depends on how teachers put these innovations into practice. Many innovations in education do not result into the desired changes. Next to institutional factors, teachers' behaviour to cope with innovations is of utmost importance for understanding the success or failure of innovations (Hasanefendic et al., 2017). Teachers play a crucial role in innovations (George & Sabapathy, 2011; Koeslag-Kreunen, Van der Klink, Van den Bossche & Gijsselaers, 2018). As Sherry (2002) states: "Individuals must be the focus if change is to be facilitated. Institutions will not change until their members change" (p.214). Innovations require that teachers develop new behaviour, but often, even after a considerable period of time, teachers abandon the newly acquired behaviour and fall back to comfortable old routines (Verloop et al., 2001). To safeguard the success of innovations it is crucial to encourage and enhance teachers' innovative behaviour (Thurlings, Evers & Vermeulen, 2015).

Innovative work behaviour (IWB) is an emerging concept, originally defined by Janssen (2000) as individual behaviour that leads to the initiation, presentation and realization of new ideas, products or procedures within the working place, team or organization. Since then this definition has been used, and further elaborated by many researchers. Also in the education context, IWB has gained attention. For example, Messmann (2012) defines IWB as a multi-stage iterative process, consisting of four different phases: 1] Opportunity Exploration which entails paying attention to trends, opportunities for innovation and problem recognition; 2] Idea Generation, i.e. generating novel and useful ideas for products, services or processes; 3] Idea Promotion, which includes seeking sponsorship for the ideas among colleagues and supervisors and apply for funding and facilitation and finally; 4] Idea Realization, which involves the creation and implementation of a prototype or model in the workgroup or organization. The phases are partly dependent, but do not necessarily follow a fixed order, resulting in a complex, dynamic, non-linear model of IWB (Messmann & Mulder, 2012). For instance, when promoting new ideas, teachers might identify new opportunities for innovation.

Though the IWB-concept has been further developed, some gaps need our attention.

Firstly, if IWB is compared with current theoretical insights on innovations, a conceptual gap can be identified. Innovation theories emphasize the need to stabilize the innovation (West & Farr, 1989). This stabilization phase is sometimes also labeled as "transfer" or "diffusion" (Kanter, 1988), or "continuation" or "institutionalization"

(Fullan, 2007). Stabilization is crucial when it comes to enhancing the sustainability of innovations. Stabilization is an extension of the implementation phase, in which the innovation is anchored into the organizational system. The integration of innovation in the organization increases its long-term success and continuity (Fullan, 2007; West & Farr, 1989). However, thus far sustainability has not been included in IWB-conceptualizations and operationalizations.

Secondly the few identified studies that have developed and empirically tested IWB-measurement instruments have been faced with challenging situations (see Table 1). The initial IWB measures consist of only a few items, representing one IWB-dimension. These measures do not reflect the complex and multidimensional nature of IWB. Other studies have explored multiple measures for distinct IWB-dimensions, but often fail to *test* the validity of these separate dimensions (Krause, 2004; Dorenbosch et al., 2005) or report limited evidence on the construct validity of the subdimensions due to four under-represented dimensions (e.g. De Jong & Den Hartog, 2010). As Messick (1995) points out: construct under-representation threatens validity. The multiple IWB measures, developed by Messmann and Mulder (2012) for the education context suffer from similar deficiencies. Their work supports the existence of multiple IWB-dimensions. However they cannot yet empirically substantiate an Idea Realization dimension.

In sum, IWB conceptualizations and operationalizations need further attention. Existing measurements miss empirical evidence of the construct validity and moreover do not include a sustainability dimension. Our study aims to develop and validate a multi-dimensional IWB instrument to measure teachers IWB. Based on a thorough and comprehensive conceptualization of IWB, we first adapted and extended the previously used instruments and developed items for a sustainability dimension. Second, we tested the construct validity of this newly developed multi-dimensional IWB instrument.

Theoretical Framework

Kanter (1988) developed the foundation for IWB, as it is known today. She suggested that the innovation process in organizations could best be understood by dividing it into tasks that individuals engage in correlating with the innovation process. The following four tasks were identified: Idea Generation, Coalition Building, Idea Realization and Transfer/Diffusion. Since Kanter, many authors have addressed the role of the individual in innovation, often referred to as innovative behaviour (e.g. Scott & Bruce, 1994; Woodman et al., 1993). Janssen (2000) proposed the term Innovative Work Behaviour for the ‘intentional creation, introduction and application of new ideas within a work role, group or organization in order to benefit role performance, the group or the organization’ (p.288). Janssen’s proposal was adopted on a wide scale (e.g. De Jong & Den Hartog, 2010; Tuominen & Toivonen, 2011; for an overview of studies in the educational context, see Thurlings et al., 2015). If we compare this elaboration with the phases of innovation proposed by Kanter, we see that the transfer / diffusion phase is not included in Janssen’s concept.

Previous Dimensions of IWB

Many scholars have discussed the multi-dimensionality of individual innovation and, later on, IWB. In 1990, Farr and Ford proposed a two-stage model that included a

Table 1 Findings from literature: Authors, number of items and dimensions, quality of measurements, number of respondents and domains

Authors	(Conceptual) Dimensions and Items	Number of Likert scale categories	Reliability and Validity	Number of Respondents	Domain
One-dimensional					
Scott & Bruce (1994)	Idea Generation, Promotion and Realization 6 items	5 point Likert scale	No validity tests reported Reliability ($\alpha = .89$).	$N = 172$	Engineers, scientists and technicians in R&D facility in industrial domain
Bunce & West (1995)	Innovative behaviour 5 items	No information available	Reliability Sample 1 $\alpha = .75$, Sample 2 $\alpha = .80$	Sample 1 $N = 435$ Sample 2 $N = 281$	Employees from a national health service
Spreitzer (1995)	Innovative behaviour 4 items	No information available	Reliability ($\alpha = .91$)	$N = 393$	Managers of an industrial company
Basu & Green (1997)	Innovative behaviour 4 items	No information available	Reliability ($\alpha = .93$), supervisor ratings.	$N = 225$	Employees of a printing manufacturer
Scott & Bruce (1998)	Innovative behaviour 4 items	7 point Likert scale	Sample 1: Reliability ($\alpha = .86$). Significant correlation with objective measure of individual innovation Sample 2: Reliability ($\alpha = .84$).	Sample 1 $N = 110$ Sample 2 $N = 149$	Sample 1: engineers and scientists in R&D facility in industrial domain, Sample 2: R&D engineers in electronic equipment.
Janssen (2000)	Idea Generation, Idea Promotion, Idea Realization 9 items	7 point Likert scale	Good reliability ($\alpha = .94$). High inter-correlations, so assumed scale to combine additively to create overall scale ($\alpha = .95$ and $\alpha = .96$ for self-ratings and supervisor ratings). Significant correlation between both scales ($r = .35$)	$N = 170$ self-ratings. $N = 110$ supervisor ratings	Employees of industrial organization in food sector.
Kleysen & Street (2001)	Opportunity Exploration, Generativity, Formative	6 point Likert scale	Reliability ($\alpha = .95$).	$N = 225$	

Table 1 (continued)

Authors	(Conceptual) Dimensions and Items	Number of Likert scale categories	Reliability and Validity	Number of Respondents	Domain
Radaelli, Lettieri, Mura & Spiller (2014) Multi-dimensional	Investigation, Championing and Application 14 items		No evidence found for 5-dimensional model		Multiple types of employees and organizations
	Innovative behaviour 4 items (based on Janssen, 2000)	7 point Likert scale	Reliability ($\alpha = .94$)	$N = 155$	Healthcare employees
Krause (2004)	(conceptual) Dimensions, and items	Number of Likert scale categories and analysis technique	Reported validity	Number of respondents	Domain
	2 dimensions Generating and testing of ideas & Implementation Total 8 items Creativity (5 items) Implementation (3 items)	7 point Likert scale Exploratory Factor Analysis (EFA)	No validity tests reported Reliability ($\alpha = .78 / .81$).	$N = 399$	Middle management of multiple types of organizations
Dorenbosch, van Engen & Verhagen (2005)	2 dimensions Creativity oriented and implementation-oriented behaviour Total 16 items Creativity (10 items) Implementation (6 items)	5 point Likert scale (EFA)	No validity tests reported Reliability ($\alpha = .88 / .90$) Conceived to combine additively as overall scale for IWB ($\alpha = .92$)	$N = 132$	Administrative employees of governmental organization
	4 dimensions 10 items Idea Exploration (2 items) Idea Generation (3 items), Idea Championing (2 items) Idea Implementation (3 items)	5 point Likert scale (EFA) and Confirmatory Factor analysis (CFA)	Reported content validity and criterion validity Empirical evidence, supporting the number of dimensions, is lacking. The author's notion that	$N = 703$	Employees and managers from small knowledge-intensive firms

Table 1 (continued)

Authors	(Conceptual) Dimensions and Items	Number of Likert scale categories	Reliability and Validity	Number of Respondents	Domain
			<p>the 10 items can be used as an overall scale make it unlikely that the 10 items represent four dimensions.</p> <p>It is likely that the four operationalizations are under-representations of the underlying variables, which threaten the construct validity (Boyle, 1991; Messick, 1995)</p> <p>Reliabilities Idea Exploration ($\alpha = .70$) Idea Exploration ($\alpha = .88$) Idea Generation ($\alpha = .90$) Idea Championing ($\alpha = .95$) Idea implementation ($\alpha = .93$)</p>		
Messmann & Mulder (2012)	4 dimensions (20 items) Opportunity Exploration (4 items) Idea Generation (6 items) Idea Promotion (7 items) Reflection (3 items)	6 point Likert scale	<p>Empirical evidence for 4 of the 5 dimensions. Idea Realization items loaded on the other factors.</p> <p>Reliabilities: Opportunity Exploration $\alpha = .76$ Idea Generation ($\alpha = .83$) Promotion ($\alpha = .87$), Reflection ($\alpha = .76$) Criterion validity and initial evidence for Content, Construct and Cross Validity</p>	$N = 335$ and $N = 293$	Employees of automotive supply organization and teachers in vocational education

creative and an implementation part. In the same vein, Krause (2004) and Dorenbosch et al. (2005) argued for a two-stage model comprising creativity –oriented and implementation –oriented work behaviour. In addition to two-stage models, multi-stage models have been proposed. For the three-stage model, the dimension Idea Promotion was added to the Idea Generation, and Idea Realization dimensions (e.g. Janssen, 2000, Scott & Bruce, 1994, 1998). In the four-stage model, scholars have included the Opportunity Exploration (OE) phase, which precedes Idea Generation. In the OE phase problems and needs in one's work context are recognized, and change opportunities are created (e.g. Janssen et al., 1997; Kleysen & Street, 2001). In a first five-stage model, Kleysen and Street (2001) added the Formative Investigation stage, referring to the Krause (2004) testing phase. This stage entails formulating new ideas and solutions, testing them, and evaluating the outcomes (Kleysen & Street, 2001). Messmann and Mulder (2012) added a fifth stage to the Opportunity Exploration, Idea Generation, Promotion, and Realization stages: the Reflection stage. This stage encompasses “assessing the progress of innovation development, evaluating activities and outcomes based on criteria for success, examining one's personal advancement during innovation development, and improving action strategies for future situations.” (p. 46)

Idea Sustainability as an Additional Dimension of IWB

The multi-stage models of IWB outlined above are, notably, comparable with innovation models described in the innovation literature. Both the innovation models of West and Farr (1989) and Fullan (2007) in the educational context propose corresponding phases. If we compare the two models, it can be observed that the first two phases of West and Farr's model (recognition and initiation) are both included in the initiation phase of Fullan's model. This initiation phase consists of change initiation, where performance gaps are recognized and new solutions or ideas are generated. Both models identify an implementation phase, during which the group executes the innovation the first time, and where effects of the implementation are observable in the work place. West and Farr completed the innovation cycle with a stabilization phase, or, as Fullan named it, a continuation phase. In this phase, the change has to become embedded in the organizational system. Hence, the continuation phase is an extension of the implementation phase, anchoring the innovation in the organization. Both models present a more encompassing definition of innovation, which is important, given that in educational settings there is a tendency to ignore the conditions (such as time) for the continuation or stabilization phase. In this context, both West and Farr, and Fullan stress the importance of a stabilization or continuation phase to sustain an innovative idea.

The prevailing conceptualizations of IWB include dimensions that refer to the first three phases of the innovation cycle (see Fig. 1). Remarkably, neither the stabilization nor the continuation phase is reflected in any of the IWB conceptualizations. Despite Kanter's (1988) emphasis on the diffusion phase, the stage has not been included in more recent conceptualizations of IWB, such as the work of Messmann and Mulder (2011), who argue that the diffusion is unrelated to innovation, with different contextual characteristics, other persons and distinct resources. However, as Van de Ven (1986) puts it: “An invention or creative idea does not become an innovation until it is implemented or institutionalized” (p.604). So, Van de Ven (1986) and Johnson et al. (2004) argue that it is necessary to focus both on the implementation of innovations in the short and the long term. It follows that those responsible for the implementation of

Fig. 1 The innovation cycle (West & Farr, 1989)

an innovation must continue to work on its sustainability, where the further continuation of the newly implemented idea must be a primary goal. This sustainability phase includes disseminating innovative ideas into the deeper structure of the organization through institutionalization (Gannaway et al., 2013; Reay et al., 2013). Particular steps have to be taken to strengthen the infrastructure that are necessary to sustain an innovative idea (Johnson, et al., 2004). The literature specifies the following features of sustainability: Improving and optimizing the innovation, such as updating and continuous regeneration to avoid implementation dips (Coffey & Horner, 2012; Fullan, 2002, Loh et al., 2013); Embedding the innovation in-depth in the system of the organization, by capacity building for securing adequate resources (Loh, et al., 2013; Fullan, 2007); Disseminating the innovation on a larger scale, like planning for project growth and broader application of an innovative idea (Loh, et al., 2013) and finally visualization of the benefits of the innovation for stakeholders by stimulating community participation and communicating a longer-term vision and outcomes (Loh, et al., 2013). In sum, we propose the following definition of IWB:

Innovative Work Behaviour is a multi-stage iterative process in which employee behaviour targets the exploration of opportunities, idea generation, idea promotion, idea realization and the sustainable implementation of these ideas, processes, products or procedures within a role, a group or an organization, whereby the ideas are (relatively) new and intended to benefit the relevant unit of adoption.

Previous Measures of IWB Excluding the Idea Sustainability Dimension

Alongside the discussions that have taken place at a conceptual level, efforts have been made to measure IWB (see Table 1). In 1994, Scott and Bruce developed the first

instrument to measure IWB ('individual innovation'). Although they define innovative behaviour as encompassing three dimensions (Idea Generation, Coalition Building and Idea Realization), their six-item instrument is a one-dimensional operationalization. Starting from a similar conceptualization (Idea Generation, Idea Promotion, Idea Realization), Janssen (2000) is the first researcher who tries to develop a multi-dimensional scale (nine items) using both self and other's ratings of IWB. Due to high correlations between the subscales he infers that the set of items represents a single scale with high reliability for the total instrument. However, both separate measures are still applied frequently, although there is insufficient information on the validity of the instruments (Thurlings et al., 2015; Yidong & Xinxin, 2013). Another attempt to develop a multi-dimensional scale is reported by Kleysen and Street (2001). They identified the dimensions: Opportunity Exploration, Generativity, Formative Investigation, Championing and Application. Their instrument has been tested in a variety of organizations, but they have not been able to provide evidence for their five-dimensional model. Studies, which have not executed validity tests, have reported a two-dimensional (Krause, 2004; Dorenbosch, et al., 2005) or four-dimensional (De Jong & Den Hartog, 2010) measure, although the latter article provides limited evidence for its construct validity. The most recent measures are provided by Messmann and Mulder (2012), who have developed measures for five dimensions namely Opportunity exploration, Idea Generation, Idea Promotion, Idea Realization and Reflection. However, the authors have not yet succeeded to identify the Idea realization dimension in the validation sample.

Table 1 illustrates the most recent attempts to develop measures for multiple IWB-dimensions. However, only a few studies also attempt to validate the measures. Merely two out of the four studies that hypothesize multiple IWB-dimensions provide empirical evidence that (partly) supports the hypothesized dimensions. The remaining two studies do not go beyond reporting initial consistency measures in terms of Cronbach's Alpha. Internal consistency is a prerequisite for validity but does not guarantee validity (Boyle, 1991). Moreover, in contrast to the original theory on IWB, the four studies fail to conceptualize and operationalize a sustainability dimension. Therefore, this study aims to develop and validate a multi-dimensional instrument for measuring IWB consisting of five dimensions, namely Opportunity Exploration, Idea Generation, Idea Promotion, Idea Realization and Idea Sustainability.

Method

Procedure and Participants

Data on the Innovative Work Behaviour scale were collected from four different vocational education institutes in the South of the Netherlands. Their board members expressed in meetings with the authors that in their institutes educational innovation is of paramount importance. All four institutes have taken initiatives to put innovation high on their agenda. Each board member has given permission to approach the Deans of the different departments in their institutes. A number of Deans responded and allowed access to the team leaders of the various education teams. These team leaders have been approached and informed by the researcher on the subject of the study.

Teacher teams were informed by team leaders and were asked to participate in the study on a voluntary basis by filling out an online questionnaire. In total, 440 teachers completed the survey. More than half of the sample was female (58.8%). On average, participants had 15.5 years of work experience in the vocational sector and the average age was 48.34 years. Respondents worked in various domains (i.e. economics, ICT, healthcare, hospitality, education etc.)

Scale Construction

Based on the literature review, we hypothesized that Innovative Work Behaviour consisted of five dimensions, namely Opportunity Exploration, Idea Generation, Idea Promotion, Idea Realization and Idea Sustainability. In turn, the scale for measuring IWB had five dimensions, namely Opportunity Exploration (OE), Idea Generation (IG), Idea Promotion (IP), Idea Realization (IR) and Idea Sustainability (IS).

The scale for measuring IWB consisted of 32 traditional IWB items (measuring OE, IG, IP and IR), based on two questionnaires developed by Messmann and Mulder (2012, 2014), completed with 20 literature-based, newly developed items for OE, IG, IP and IR and the sustainability dimension. All items were subjected to reflective discussions between the researchers, keeping in mind the original theoretical framework on IWB, and the broader conceptualization of educational innovations as discussed above. The pool of 52 items was supposed to represent the right set of items to cover the five dimensions. If needed, items were slightly adapted and reformulated to better attune the target group of vocational teachers. This was the case with two OE items (e.g. *exchanging thoughts on recent developments or problems at work with one's colleagues*), three IG items (e.g. *discussing personal suggestions for improvements with one's colleagues*), one IP item (e.g. *providing insight on the step-by-step transformation of the new idea into practice*) and two IR items (e.g. *Drawing up possible operational strategies for future and comparable situations*). Finally, two new items were added relating to IR, to better discriminate between the dimensions idea promotion and idea realization (e.g. *Supporting colleagues with the application of a developed idea*). For the development and the customization of the items, we followed the steps for scale construction as proposed by Spector (1992). The content validity was assessed by verifying whether the literature-based operationalizations genuinely represented the intended content domain for each construct (Boyle, 1991; Curda, 1997; Linacre, 2006; Trochim, 2002).

Following Spector's (1992) steps for scale construction once again, twenty new items were developed as operationalizations of the sustainability dimension. The following search terms were used in the preceding literature review: *sustainability, durability, transfer, adoption, diffusion, continuation of innovations, dissemination of innovations, institutionalization, maintaining innovations and scalability of innovations*, both in English and American orthography. Second, results (definitions of sustainability, and its features) from the literature search, were presented to experts ($N=4$) in the field of educational development (a professor corporate learning in business and economics; a professor in educational sciences, and two senior researchers in educational psychology and learning and development). Collectively, the most relevant definition and features of sustainability were determined, namely improving and optimizing the innovation, disseminating the innovation on a larger scale, embedding the innovation in-depth in the system of the organization, and visualizing the

benefits of the innovation for stakeholders. Third, 20 items were developed in accordance with these features. The constructed items were presented to the same expert group to judge their relevance and clarity. Fourth and final, a pilot study was employed, in which the questionnaire with all five dimensions was presented to a group of teachers, who were active members of an educational research community for continuous professional development in an institute for vocational education in the Netherlands ($N=10$). The participants in the pilot study matched the target group of vocational teachers. Feedback on the items was processed, and in a second session, the questionnaire was scrutinized for clarity, layout and the degree to which the items subjectively appeared to measure the constructs (face validity). The pilot study resulted in minor adaptations in wording of the items and layout of the questionnaire. Furthermore the pilot study revealed that the respondents only used a restricted number of answer options. The person response patterns deviated from the intended six-point scales (1 = does not apply, 6 = fully applies). The rating scales were not adapted, due to the small size of the pilotgroup, but it led to the choice of the Rasch rating scale model (Rasch, 1960) to analyze the scale structures of the five intended dimensions. All scales are introduced by the sentence: “to what extent do the following work activities apply to you”. The Appendix lists all IWB items (52 items).

The questionnaire also included items that measured personal background variables, such as gender, age, level of education, job tenure, working hours and job position. All these characteristics have shown to be significantly related to IWB (e.g. Baer et al., 2003; Janssen, 2005; Messmann & Mulder, 2014; Pieterse et al., 2010).

Data Analysis

Preliminary Analysis

The analysis was conducted in two stages. The preliminary analysis involved scrutinizing the IWB scales with the Rasch rating scale model (Rasch, 1960) to create invariant interval measures and to evaluate the construct validity of the IWB dimensions. Also, to shed more light on the restricted person responses patterns on the items of the scale, as observed in the pilot group. The IWB survey (including its new scale Sustainability) employs six-point Likert scales. Likert scales consists of ordered (ordinal) raw scores. A 6-point Likert scale consists of six ordered categories, where the second category in a scale item represents more of the attribute than the first category, the third more than the second, and so on. However it is unknown whether the psychological distances between each of the categories are equal to each other (e.g. is the distance from strongly agree to agree equal to the distance between agree and disagree). On top of that, it is unknown whether respondents can genuinely distinguish six substantive differences in a scale item. The fact that the response format is providing six response options for each item, does not guarantee that all six categories are genuinely used by participants. A Rasch model can diagnose how many categories are distinguished by persons in items and across sets of items. Subsequently, the Rasch model can transform ordered raw scores into equal measurement units (interval level) if the data fit the Rasch model. Like a ruler or a thermometer, a Rasch measure is composed of equal, interval measurement units, which guarantees invariant measurements across samples. Also, Rasch models allow researchers to examine person and item measures simultaneously on the same interval scale.

In our study scales and items were developed based on previous research and content related literature. Rasch analyses were used as confirmatory tests of the extent to which scales have been successfully developed according to prior measurement criteria (Ludlow et al., 2008).

For the present study, the Rasch Rating Scale Model was used (Rasch, 1960). The Rasch analyses were executed in WINSTEPS 4.0, which uses joint maximum likelihood estimation (JMLE; Linacre, 2017).

The transformation of ordered qualitative observations into additive measures is a Rasch model. Rasch models are logit-linear models. The polytomous Rasch rating scale model that is applied in the current study uses the following additive transformation (Linacre, 2017, p. 34):

$$\log(P_{nij}/P_{ni}(j-1)) = B_n - D_i - F_j$$

Where P_{nij} is the probability that person n encountering item i is observed in category j ; B_n is the ability of person n ; D_i is the difficulty measure of item i , the point where the highest and lowest categories of the item are equally probable. F_j is the calibration measure of category j relative to category $j-1$, the point where categories $j-1$ and j are equally probable relative to the measure of the item. No constraints are placed on the possible values of F_j .

To scrutinize the IWB scales, six requirements were investigated (e.g. Wolfe and Smith, 2007): Rating scale effectiveness, dimensionality, reliability, item measure quality, person measure quality and item hierarchy.

Confirmatory Factor Analysis

The second phase of the analyses consisted of confirmatory factor analysis (CFA), with the results of the Rasch analysis, in AMOS, version 25 (Arbuckle, 2006) to analyse the structural relationship between the items and their latent variables and the intercorrelations between the variables. To assess the model fit, a variety of fit indices were used, to be able to illuminate different aspects of goodness of fit (Lomax and Schumacker, 2004). As recommended by Schermelleh-Engel et al. (2003), we used the Chi-square test (X^2), a likelihood ratio test statistic. The chi-square statistic postulates that the specified factor loadings, factor variances, factor covariances and error variances are valid. The chi-square/df ratio (X^2/df), provides information on model parsimony. A rule of thumb is that X^2/df values between 1 and 3 are indicative for parsimonious models (Tabachnick et al., 2007). The Tucker-Lewis Index (TLI) informs us of the adequacy of a nested model. A TLI > .90 represents a good fit (Byrne, 2016). The Comparative Fit Index (CFI; Bentler, 1990) also compares the specified model with the baseline model. A CFI > .95 represents an excellent fit (Byrne, 2016; Hu & Bentler, 1999). Finally the Root Mean Square Error of Approximation (RMSEA; Hair et al., 2010) evaluates the model fit, taking into account the complexity of the model. RMSEA values < .05 indicate good fit (Hair et al., 2010).

Results

First the process of the scrutinizing the IWB-scales in the Rasch model, based on the six criteria, is described, followed by the confirmatory factor analysis in AMOS-25.

Rasch Analyses

Rating scale effectiveness indicates how well each scale is functioning and, how well each of the IWB scales fits the Rasch model. As described previously, a Rasch model allows us to diagnose how many categories are distinguished by persons, in sets of items. Subsequently, if a (recategorized) scale fits the Rasch model, the ordered raw scores can be transformed in a scale that consists of equal measurement units (interval level). These Rasch interval measures are invariant across samples, and allow researchers to examine person measures and item measures simultaneously on the same interval scale. The rating scale effectiveness was evaluated by analysing each scale's average item INFIT and OUTFIT mean square fit statistics. In survey development researchers mostly focus on item construction. Boone et al., 2014 recommended that for items, the item outfit mean square values (MNSQ's) should be inspected. Outfit MNSQ's are sensitive to outliers in the data; infit MNSQ's are sensitive to unexpected behaviour close to the item's difficulty (or person's endorse ability) level. Outfit MNSQ's, as well as infit MNSQ's of attitudinal data should be $\geq .50$ and ≤ 1.5 .

Mean square values ≥ 1.5 indicate noise in the data; $\leq .50$ suggest dependency. Table 2 shows that the average item outfit mean squares are within the range of .92–1.21 logits, which demonstrates that the scale measures meet the requirements of the Rasch model.

As described previously, the Rasch model allows us to scrutinize item measures and person measures simultaneously, and on the same interval scale. The person measure quality is assessed by evaluating each scale's average person measure (the person Mean) and by evaluating each scale's person infit and outfit mean square value (Table 2). A scale's Person Mean reflects the degree to which a scale is tuned to the target group. The person Means, which are recalibrated into their (recalibrated) Likert scores vary from 3.80 ($SD = 0.75$) to 5.07 ($SD = 0.60$). High scale person Means are indicative for scales that are too easy to endorse by the target group (e.g. the creativity constructs OE, IG and IP). In addition, also the person standard deviation should be taken into account, to diagnose how persons are dispersed along the latent variable of each scale. Table 2 demonstrates that the easiest scale to endorse IP, also shows the smallest spread along the underlying variable. Similar to the item fit requirements, the average person infit and outfit MNSQ's should be $\geq .50$ and ≤ 1.5 (Linacre and Wright, 1994).

Dimensionality is assessed by principal components analysis of the standardized residuals of each (restructured) scale's residuals. This analysis identifies how much variance is explained by the person and the item measures. The identification of Eigen Values >2 indicate that two or more items might (also) measure something different than the latent variable. This can then be verified by a statistical comparison of the person measures of both possible dimensions. Disattenuated correlations (correlations, corrected for measurement errors) between the possible dimensions $\geq r = .70$ that co-occur with $\leq 5\%$ (max $<10\%$) of the person measures outside the 95% confidence boundaries are strong indicators that a construct is one-dimensional (Linacre, 1998). If more than the expected 5% (max 10%) of the person responses falls outside the confidence boundaries the likelihood increases that more than one dimension is measured. Two scales showed to be two-dimensional. Idea Realization differentiated into two subscales and also Idea Sustainability was divided into two different subscales (see Table 2).

Table 2 Rating scale effectiveness

	Measure	Model Error	INFIT MNSQ	OUTFIT MNSQ
Dimension Opportunity Exploration				
Items ($n = 4$) item structure 4–5–6				
Item Mean	.00 (.47)	.12	.99	1.00
Person Mean	5.07 (0.61)	1.33	.99	1.00
Dimension Idea Generation				
Items ($n = 7$) item structure 4–5–6				
Item Mean	.00 (.25)	.12	.99	.98
Person Mean	5.07 (0.61)	1.16	.98	.98
Dimension Idea Promotion				
Items ($n = 7$) item structure 3–4–5–6				
Item Mean	.00 (.47)	.10	.99	.99
Person Mean	5.08 (0.60)	.94	.98	.98
Dimension Idea Realisation (Criterion-Based implementation)				
Items ($n = 4$) item structure 2–3–4–5–6				
Item Mean	.00 (.31)	.09	.99	.98
Person Mean	3.84 (1.01)	1.15	1.20	1.20
Dimension Idea Realisation (Learning-Based communication)				
Items ($n = 5$) item structure 3–4–5–6				
Item Mean	.00 (.94)	.08	1.00	.96
Person Mean	4.23 (0.68)	.96	.97	.96
Dimension Idea Sustainability (External Dissemination)				
Items ($n = 6$) item structure 3–5				
Item Mean	.00 (.65)	.10	.99	1.21
Person Mean	3.80 (0.75)	.99	.97	1.20
Dimension Idea Sustainability (Internal embedding)				
Items ($n = 11$) item structure 3–5				
Item Mean	.00 (.67)	.08	1.02	.92
Person Mean	4.10 (0.63)	.73	1.00	.92

Means of Rasch items (thus also the Rasch scale means) are always arbitrarily fixed to zero. The presented scale means are average person measures. The average person means and person standard deviations (in brackets, after the person Mean) are recalibrated into the original raw Likert scores. The remaining fit statistics are expressed in logits

Person Reliability and *Person strata* indicate the extent to which scores are reproducible and/or the extent to which a scale can distinguish endorse ability differences between persons (see Table 3).

Person reliability is expressed in Cronbach's alpha and refers to the internal consistency of a scale. Cronbach's alphas vary from .84 to .94 indicating good to excellent internal consistency for each scale. *Person strata* represent how many different endorse ability levels between persons can be distinguished which cannot be attributed to measurement errors. Person strata should at least be ≥ 2.0 . All person strata meet this criterion except for Idea sustainability External Dissemination (1.89).

Table 3 Reliability and Strata measures ($N = 440$)

Dimension	Reliability	Strata
Opportunity Exploration (OE)		
Persons	.84	2.51
Items	.94	5.63
Idea Generation (IG)		
Persons	.94	3.48
Items	.76	2.68
Idea Promotion (IP)		
Persons	.94	4.19
Items	.96	6.71
Idea Realization Criterion-Based Implementation (CBI)		
Persons	.93	4.21
Items	.91	4.45
Idea Realization Learning-Based Communication (LBC)		
Persons	.87	3.08
Items	.99	15.37
Idea Sustainability External dissemination (ISED)		
Persons	.90	1.89
Items	.98	9.28
Idea Sustainability Internal Embedding (ISIE)		
Persons	.92	2.49
Items	.99	11.32

The person reliability is expressed in terms of Cronbach's alpha; Item reliability is a Rasch specific item-difficulty-order reproducibility statistic

Item Reliability and Item strata indicate the extent to which scores are reproducible and/or the extent to which a scale can distinguish difficulty differences across items. *Rasch item reliability* is a Rasch characteristic specification of the item-difficulty-order reproducibility. Rasch item reliabilities range from .76 to .99, indicating good to almost perfect item reliability. *Item strata* represent how many distinct difficulty levels between items can be distinguished, which cannot be attributed to measurement errors. Item strata should at least be ≥ 2.0 . Item strata vary from 2.68 to 15.37. All item strata meet this criterion (Table 3).

Item measure quality is evaluated by examining the item hierarchy, the extent to which the items vary in difficulty, the size of the standard errors, and the degree to which the items fit the expectation of the Rasch model (see Table 4). *Item hierarchy* implies that items should be ordered from bottom to top, from easiest to endorse to most difficult to endorse. Items rank-order themselves in a manner that is consistent with the theory. The item hierarchical maps of all subdimensions are investigated and were hierarchically ordered (see Table 4).

Item difficulty calibrations encompassed ranges from -0.30 to 0.40 logits (OE) to -1.69 to 1.08 logits (IRLBC), indicating a weak to a good distribution of items along their latent variable (Table 4). Standard errors are small, ranging from 0.08 to 0.12

Table 4 Item measure quality, as demonstrated by item measure hierarchical order, item measurement errors, and the outfit mean square statistics

Item	Measure	SE	INFIT MNSQ	OUTFIT MNSQ
Dimension Opportunity Exploration				
OE7	.74	.11	.89	.89
OE8	-.01	.11	.85	.85
OE6	-.16	.11	.76	.76
OE5	-.57	.11	1.47	1.50
Dimension Idea Generation				
IG7	.40	.12	.77	.72
IG5	.29	.12	.86	.88
IG3	.08	.12	1.20	1.17
IG1	-.09	.12	1.31	1.36
IG4	-.16	.12	.96	.92
IG6	-.22	.12	.87	.86
IG2	-.30	.12	.98	.95
Dimension Idea Promotion				
IP7	.91	.09	1.08	1.09
IP2	.23	.09	1.03	1.00
IP1	.20	.09	1.04	1.07
IP3	-.16	.09	1.06	1.04
IP6	-.20	.09	.93	.92
IP5	-.31	.09	.99	.97
IP4	-.67	.10	.81	.82
Dimension Idea Realisation (Criterion-Based implementation)				
IR5	.37	.09	1.17	1.17
IR2	.13	.09	1.06	1.06
IR3	-.03	.09	.76	.75
IR4	-.47	.09	.96	.94
Dimension Idea Realisation (Learning-Based communication)				
IR6	1.08	.08	1.10	1.07
IR9	.54	.08	1.00	.94
IR7	.24	.08	.81	.78
IR10	-.17	.08	1.09	1.06
IR8	-1.69	.08	1.00	.96
Dimension Idea Sustainability (External Dissemination)				
ISV4	1.04	.11	1.13	1.89
ISV3	.58	.09	1.18	1.67
ISDL4	.08	.09	.77	.71
ISDL3	-.31	.09	.83	.84
ISDL1	-.49	.09	.94	1.05
ISDL2	-.90	.09	1.06	1.13
Dimension Idea Sustainability (Internal embedding)				
ISV1	1.47	.09	1.16	.97

Table 4 (continued)

Item	Measure	SE	INFIT MNSQ	OUTFIT MNSQ
ISDD4	.82	.08	1.01	.90
ISV2	.59	.08	1.07	.95
ISDD3	.24	.08	.89	.85
ISIO6	-.02	.08	1.00	1.14
ISIO2	-.31	.08	1.07	1.03
ISIO1	-.45	.08	1.09	.92
ISIO7	-.46	.08	1.02	.91
ISDD2	-.49	.08	.99	.86
ISDD1	-.66	.08	1.05	1.11
ISIO8	-.73	.08	.82	.54

Rasch measures are invariant interval measures, within the boundaries of their measurement errors. Rasch measures are expressed in logits. All item calibrated measures are already hierarchically ordered from easiest to endorse (at the bottom of the scale) to most difficult to endorse (top)

logits. OUTFIT mean square values, which are sensitive to outliers, are the most important quality indicators for items (Boone et al., 2014). Linacre and Wright (1994) suggest that outfit MNSQ's as well as infit MNSQ's for attitudinal data should preferably be ≥ 0.5 and ≤ 1.5 ideal. Only item ISV4 and ISV3 (dimension Idea Sustainability external dissemination), show some noise in their OUTFIT mean square statistics (1.89 and 1.67 respectively).

Eight items out of the pool of 52 items were deleted because they did not meet the requirements of the Rasch Rating Scale model, neither in the original scale structure format, nor the restructured one (see appendix).

Person Background Variables and the IWB-Dimensions

As described in the scale construction section, our questionnaire also encompassed personal background variables. Previous research has reported significant relationships between IWB and personal background variables. The Rasch model has not only provided valuable information on how IWB items should be restructured in order to produce invariant person and item measures for all hypothesized IWB scales (see Table 2), but also allows us to utilize these measures to get an impression of the relationships between person background characteristics and the distinct IWB-dimensions. Exclusively the significant relationships are now described below.

Significant relations between person background characteristics and IWB are differentiated across the distinct dimensions. Age correlates significantly and negatively with Idea Sustainability external dissemination, $r = -.134$, $p < .01$. Significant positive correlations between being a woman and the creative IWB-dimensions are found: OE, $r = .135$, $p < .01$, IG, $r = .128$, $p < .01$, and IP, $r = .100$, $p < .05$, supplemented by a significant positive correlation between being a woman and IR Learning Based Communication, $r = .128$, $p < .01$. A higher previous education level is significantly and positively related to all Idea Realization and Idea sustainability dimensions. IRCBI, $r = .126$, $p < .01$, IRLBC, $r = .129$, $p < .01$, ISED, $r = .217$, $p < .01$ and, ISIE, $r = .180$,

$p < .01$. Work hours ratio is positively and significantly related to IP, $r = .105$, $p < .05$, IRCBI, $r = .133$, $p < .01$, and both Idea sustainability dimensions (respectively $r = .138$, $p < .01$, for ISED and $r = .176$, $p < .01$, for ISIE). To complete with, being employed in Vocational Education (college level), as compared to being employed in a University of Applied Science is significantly, and negatively correlated with IP, $r = -.150$, $p < .01$, IRCBI, $r = -.178$, $p < .01$, IRLBC, $r = -.229$, $p < .01$, ISED, $r = -.286$, $p < .01$ and ISIE, $r = -.268$, $p < .01$.

CFA

Most researchers are more familiar with confirmatory factor analysis (CFA) than the Rasch model. Hence, to demonstrate the construct validity, the extent to which a set of measured items actually reflects the theoretical latent construct, we test the relationships between items and their latent variables (i.e., the path estimates linking the underlying variable to indicator variables) in CFA. In CFA higher loadings confirm that the indicators (items) are strongly related to the latent variable. Significant loadings confirm construct validity. A rule of thumb suggests that the loadings should be at least .5 and ideally .7 or higher. Lower loadings suggest that an item is a candidate for deletion from the model (Hair et al., 2010). Table 5 shows that all loadings are between .59–.90 which indicate good construct validity. Item means scores and standard deviations are also displayed in Table 5.

Further we tested the seven-factor model, containing the following dimensions: Opportunity Exploration, Idea Generation, Idea Promotion, Idea Realization (both Criterion Based implementation and Learning based communication) and the dimension Idea Sustainability (both Internal Embedding and External Dissemination). This model showed a good fit ($\chi^2 = 1658$, $df = 864$, $\chi^2/df = 1.92$, CFI = .95, TLI = .94 and RMSEA = .046).

General Discussion

The urgent needs for innovation as well as the problems associated with the rapid pace of innovations have put the concept of IWB on the educational research agenda (Thurlings, et al., 2015). However, the lack of cross-validated measurement instruments, which conceptually fit the literature on innovation (cycles), jeopardizes our understanding about why innovations in educational institutions fail or succeed. Our review of the IWB measurement research has shown that over the last three decades several IWB measures have been developed, mainly based on the work of Scott and Bruce (1994). However, the studies, conducted in a variety of settings show mixed psychometric results as well as differences in the number of dimensions, due to a lack of verifying the content validity, and testing the construct validity (Table 1) (Boyle, 1991; Messick, 1995).

Therefore, the purpose of this study was to develop and validate an instrument to measure innovative work behaviour of teachers in vocational education, covering all aspects of IWB, including sustainability, a dimension that was neglected in previous studies. We have hypothesized that IWB consists of five dimensions, namely OE, IG,

Table 5 CFA results Items, means, standard deviations, and factor loadings of the Innovative work behaviour Instrument

Item wording			Factor							
		<i>M</i>	<i>SD</i>	1	2	3	4a	4b	5a	5b
<i>Opportunity Exploration (OE)</i>										
1	Questioning the current concepts, work processes and results with the goal to improve them#	5.08	.76	.83						
2	Discussing the possible leeway for change with colleagues#	4.94	.76	.80						
3	Questioning the effectiveness of the current way of working#	5.10	.76	.83						
4	Exchanging thoughts on recent developments or problems at work with colleagues#	5.17	.78	.59						
<i>Idea Generation (IG)</i>										
8	Asking critical questions about current situations at work#	5.09	.75		.72					
9	Suggesting improvements on expressed ideas^	5.01	.78		.89					
10	Exchanging ideas on concrete changes at work with one's colleagues #	5.10	.75		.85					
11	Specifying which elementary improvements can be implemented at work#	5.02	.76		.87					
8	Discuss personal ideas for improvement with one's colleagues#	5.09	.76		.84					
9	Expressing a personal opinion of underlying problems in the workplace#	5.06	.78		.76					
11	Suggesting new ideas to solve problems in the current work situation#	5.12	.75		.82					
<i>Idea Promotion (IP)</i>										
12	Suggesting the new idea to key persons who are authorized to allocate resources for this new idea^	4.59	.99			.81				
13	Convincing others of the importance of a newly developed idea or solution^	4.60	.94			.81				
14	Promoting new ideas to the supervisor in order to gain his/her active support^	4.69	1.00			.82				
15	Promoting new ideas to colleagues in order to gain their active support^	4.81	.92			.87				
16	Promoting the application of a new solution within one's work context^	4.73	.93			.84				
17	Getting colleagues acquainted with the utilization of the new idea or solution#	4.70	.93			.82				
18	Making it clear to others, how a new idea can be stepwise put into practice#	4.43	.96			.79				
<i>Idea Realization (Criterion-Based Implementation, IRCBI)</i>										
19	Defining criteria of success for the realization of the idea^	3.81	1.19				.85			
20	Monitoring the progress during the process of putting ideas in practice^	4.03	1.19				.90			
21		3.92	1.22				.88			

Table 5 (continued)

Item wording			Factor							
	<i>M</i>	<i>SD</i>	1	2	3	4a	4b	5a	5b	
Analyzing the solutions that are found on undesired effects, when putting ideas into practice^										
22 Testing solutions for unexpected problems that emerge, when putting ideas into practice^	3.87	1.21					.82			
<i>Idea Realization (Learning-Based Communication, IRLBC)</i>										
23 Obtaining information from people who have already put the idea into practice, about possible bottlenecks during the implementation process+	4.27	1.02						.72		
24 Designing operational strategies for future, comparable situations#	4.03	.99						.76		
25 Reflecting critically on the actions you execute when putting the idea into practice#	4.80	.97						.73		
26 Reflecting systematically on your experiences when putting the new idea into practice#	4.14	1.04						.78		
27 Keeping colleagues posted about the progress of the realization of the idea^	3.86	1.01						.75		
<i>Idea Sustainability (External Dissemination, ISED)</i>										
28 Discussing broader applications of the implemented idea with colleagues outside your team+	3.94	1.00							.81	
29 Participating in networks that have the innovation or new idea as a theme+	4.05	1.00							.79	
30 Initiating collaboration with other groups in the organization to apply the idea in other contexts also+	3.88	.99							.80	
31 Initiating collaboration with other groups outside of the organization to apply the idea in other contexts also+	3.76	.97							.80	
32 Communicating explicitly the returns of the implemented idea outside the team+	3.62	.93							.69	
33 Visualizing the output of the implemented ideas to a broader audience+	3.52	.88							.63	
<i>Idea Sustainability (Internal Embedding, ISIE)</i>										
34 Exchanging information about bottlenecks with people who have already implemented the idea+	4.33	.94							.71	
35 Comparing the results of the implemented idea with the predetermined, original goals+	4.27	.96							.68	
36 Initiating quality assurance systems that support the implemented idea+	3.82	.98							.67	
37 Being aware of the steps that can be made to make a success of the implementation process+	4.34	.94							.70	
38 Communicating explicitly the returns of the implemented idea, in the team+	3.91	1.00							.68	
39 Executing improvement activities to optimize the implemented solutions+	4.43	.91							.70	
40 Organizing activities for professional development for yourself and your colleagues, to continue the development of the idea+	4.16	.99							.75	

Table 5 (continued)

Item wording			Factor							
		<i>M</i>	<i>SD</i>	1	2	3	4a	4b	5a	5b
41	Actively gathering results of the implemented ideas or solutions+	4.33	.95							.66
42	Showing initiative to anchor the new idea in existing procedures or structures of the organization+	4.05	1.00							.74
43	Discussing with colleagues how implemented ideas can be embedded more firmly in the system of the organization+	4.40	.92							.69
44	Substantiating the implemented ideas with figures+	3.61	.92							.59

Note: Items were translated into English. Scale structures were adapted based on Rasch analyses and transformed from six-point Likert scales to adapted scales. Scale structures: OE 3 point scale (4,5,6), IG 3 point scale (4,5,6), IP 4 point scale (3,4,5,6), IRCBI 5 point scale (2,3,4,5,6), IRLBC 4 point scale (3,4,5,6), ISED 2 point scale (3,5) ISIE 2 point scale (3,5)

^ Original item, Messmann and Mulder (2012) validated study

Adapted items inspired by Messmann and Mulder (2014) and oral communication, no validation study available

+ Newly developed item

IP, IR and IS. Preliminary Rasch analyses and CFA, have confirmed our hypothesis that IWB consists of these five main dimensions. Additional Rasch dimensionality analyses have differentiated both the realization and sustainability dimension into two sub-dimensions (e.g. Idea Realization Criterion-Based implementation and Learning Based communication and Idea Sustainability External Dissemination and Internal Embedding). The differentiation of the IR dimension into an IR- Criterion Based implementation and IR-Learning Based Communication dimension is in line with the study of Messmann (2012), who distinguished physical and cognitive activities, considering reflection as a crucial phase in the innovative process. In our research both physical and cognitive activities are represented in the two sub-dimensions. Also, scale structures have been adjusted to improve the scale effectiveness and to guarantee that the measures are invariant across samples. In the Rasch model we have meticulously scrutinized the six requirements, as proposed by Wolfe and Smith (2007): Rating scale effectiveness, dimensionality, reliability, item measure quality, person measure quality and item hierarchy.

The model fit of the available ordinal scores on the Likert scales had to be examined in the Rasch model. Items that did not meet the requirements of the Rasch model were removed, and, if necessary, scale structures were adjusted to improve the scale's measurement quality and to assure invariance of measurement. The scale structures developed and tested in the Rasch model, were tested with CFA. This seven factor model (containing OE, IG, IP, IR with two subdimensions and IS with two subdimensions) showed a good fit.

With respect to the newly developed sustainability dimension, the identified, literature-based main features reflecting sustainability are represented in both sub-dimensions. The items reflecting *Improving and optimizing the innovation*, such as updating and continuous regeneration (Coffey & Horner, 2012; Fullan, 2002, Loh

et al., 2013) and *embedding the innovation in depth in the system* of the organization (Loh, et al., 2013; Fullan, 2007) are clustered in the sub-dimension Internal embedding. In practice we observe that the internal embedding (the implementation) is more successful when there are meetings with teachers before and after the implementation. Such meetings make it possible to discuss the training needs of the teachers involved, and to involve them in any adjustments they deem necessary for the implementation to succeed. This approach reduces the likelihood that teachers, fall back on their former behaviour (Wolbers et al., 2017) and it emphasize the importance to invest in time after the realization phase.

Items reflecting *Disseminating the innovation on a larger scale*, like planning for project growth (Loh, et al., 2013) and *visualization of the benefits* of the innovation for stakeholders (Loh, et al., 2013) are clustered in the sub-dimension External Dissemination. By adding the sustainability dimension to the current conceptualization of IWB, we are strongly in line with the innovation cycles of West and Farr (1989) and Fullan (2007), who emphasized the importance of a stabilization or continuation phase as a vital stage to complete an innovation process. The emphasis on this sustainability phase can help schools to firmly anchor innovative ideas in their organizations, preventing that time and energy are wasted through unfinished innovations.

The newly developed and validated IWB scale may help researchers to empirically examine this phenomenon more accurately, making it possible to gain more knowledge about its antecedents and consequences.

Limitations and Future Research

This study has several limitations that should be addressed in future research and questionnaire validity testing. Firstly, even though we used data from a heterogeneous sample including teachers from different professional institutes, the results confirm a five-structure model of IWB in the vocational educational sector only. Hence, future studies might benefit from the validation of our instrument by utilizing the measures and operationalizations in samples, consisting of professional and general college level and/or university level education. Secondly, our study only included Dutch participants. The obtained Rasch measures are invariant, but we have not yet explored internationally composed samples. The English version, which is provided in this paper (see appendix, and Table 5), will also allow other researchers to use these validated measures in other countries. Third, we have only reported self-ratings of the IWB measures. Future research could also investigate the perception of the direct manager, for example. However that would require a revision of the questionnaire and thus another validation. Finally, the newly developed, theory-based instrument, with its new sustainability dimension, needs to be further developed. Although it proved to be a reliable and valid instrument, items can be added to increasingly better represent the experiences with the innovation cycle in practice. Such additional items will enhance the robustness of the operationalization of the theoretical concepts of IWB.

By providing a reliable and valid scale at the interval level to measure innovative work behaviour, we hope to stimulate and encourage other researchers in the field to engage in research aimed at understanding why innovations fail or succeed. Collecting qualitative data (e.g. focus group or interviews with teachers) could help to deepen our understanding of the found quantitative data. Focus on the sustainability phase may

help to prevent superficial or incomplete implementations. Although a great deal of literature has addressed the importance of the role of individuals in making innovations happen, many questions remain unanswered on how to support individuals in taking up this role and how to manage sustainable innovations. The availability of the IWB scales, supplemented with a sustainability dimension, is the first step towards further development of a model that identifies the antecedents of IWB.

Appendix

Original item pool to measure Innovative Work Behaviour.

* item dropped after Rasch analysis.

All scales are introduced by the sentence: “to what extent do the following work activities apply to you”?

Opportunity Exploration

Keeping yourself posted about the organizations’ general structures and processes. *.

Keeping yourself posted about the latest developments in the organization. *.

Keeping yourself informed about new concepts and insights in your professional field. *.

Keeping yourself informed about new developments in other organizations *.

Exchanging thoughts on recent developments or problems at work with colleagues.

Questioning the effectiveness of the current way of working.

Discussing the possible leeway for change with colleagues.

Questioning the current concepts, work processes and results with the goal to improve them.

Idea Generation

Asking critical questions about current situations at work.

Suggesting new ideas to solve problems in the current work situation.

Expressing a personal opinion of underlying problems in the workplace.

Discuss personal ideas for improvement with one’s colleagues.

Specifying which elementary improvements can be implemented at work.

Exchanging ideas on concrete changes at work with one’s colleagues.

Suggesting improvements on expressed ideas.

Idea Promotion

Convincing others of the importance of a newly developed idea or solution.

Suggesting the new idea to key persons who are authorized to allocate resources for this new idea.

Promoting new ideas to the supervisor in order to gain his/her active support.

Promoting new ideas to colleagues in order to gain their active support.

Promoting the application of a new solution within one’s work context.

Getting colleagues acquainted with the utilization of the new idea or solution.

Making it clear to others, how a new idea can be stepwise put into practice.

Idea Realization

Supporting colleagues with the application of an already developed idea*.

Testing solutions for unexpected problems that emerge, when putting ideas into practice.

Analyzing the solutions that are found on undesired effects, when putting ideas into practice.

Monitoring the progress during the process of putting ideas into practice.

Defining criteria of success for the realization of the new idea.

Keeping colleagues posted about the progress of the realization of the idea.

Reflecting systematically on your experiences, when putting the new idea into practice.

Reflecting critically on the actions you execute, when putting the idea into practice.

Designing operational strategies for future, comparable situations.

Obtaining information from people who have already put the idea into practice, about possible bottlenecks during the implementation process.

Idea Sustainability

Actively gathering results of the implemented ideas or solutions.

Comparing the results of the implemented ideas with the predetermined, original goals.

Asking directly involved colleagues for improvements of the ideas that have been put into practice*.

Discussing possible crucial improvements to the solution that has been put into practice with your supervisor*.

Suggesting improvements to your colleagues or supervisor, after a meticulous analysis of the results of the solution*.

Organizing activities for professional development for yourself and your colleagues, to continue the development of the idea.

Exchanging information about bottlenecks with people who have already implemented the idea.

Executing improvement activities to optimize the implemented solutions.

Discussing broader applications of the implemented idea with colleagues outside your team.

Participating in networks that have the innovation or new idea as a theme.

Initiating collaboration with other groups in the organization to apply the new idea in other contexts also.

Initiating collaboration with other groups outside of the organization to apply the new idea in other contexts also.

Discussing with colleagues how implemented ideas can be embedded more firmly, in the system of the organization.

Being aware of the steps that can be made to make a success of the implementation process.

Showing initiative to anchor the new idea in existing procedures or structures of the organization.

- Initiating quality assurance systems that support the implementation of the idea.
- Substantiating the implemented ideas with figures.
- Communicating explicitly the returns of the implemented idea, in the team.
- Communicating explicitly the returns of the implemented idea, outside the team.
- Visualizing the output of the implemented innovation to a broader audience.

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

- Arbuckle, J. L. (2006). IBM® SPSS® Amos™ 25 User's guide.
- Baer, M., Oldham, G., & Cummings, A. (2003). Rewarding creativity: When does it really matter? *Leadership Quarterly*, 14, 569–586.
- Basu, R., & Green, S. G. (1997). Leader-member exchange and transformational leadership: An empirical examination of innovative behaviours in leader-member dyads. *Journal of Applied Social Psychology*, 27(6), 477–499.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Understanding person measures. In *Rasch analysis in the human sciences* (pp. 69–92). Springer, Dordrecht.
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 12(3), 291–294.
- Bourgonjon, J., De Grove, F., De Smet, C., Van Looy, J., Soetaert, R., & Valcke, (2013). Acceptance of game-based learning by secondary school teachers. *Computers & Education*, 67, 21–35.
- Bunce, D., & West, M. A. (1995). Self perceptions and perceptions of group climate as predictors of individual innovation at work. *Applied Psychology*, 44(3), 199–215.
- Byrne, B. M. (2016). Structural equation modeling with AMOS: Basic concepts, applications, and programming. Routledge.
- Coffey, J. H., & Horner, R. H. (2012). The sustainability of Schoolwide Positive Behaviour Interventions and Supports. *Council for Exceptional Children*, 78(4), 407–422.
- Curda, L. K. (1997). *Validation of a measure of teacher's self-efficacy and outcome expectations in the content domains of reading and mathematics*. Norman: University of Oklahoma.
- De Jong, J., & Den Hartog, D. (2010). Measuring innovative work behaviour. *Creativity and Innovation Management*, 19(1), 23–36.
- Dorenbosch, L. W., Van Engen, M. L., & Verhagen, M. (2005). On-the-job innovation: The impact of job design and human resource management through product ownership. *Creativity and Innovation Management*, 14(2), 129–141.
- Eteokleous, N. (2008). Evaluating computer technology integration in a centralized school system. *Computers & Education*, 51(2), 669–686.
- Farr, J., & Ford, C. (1990). Individual innovation. In M. West & J. Farr (Eds.), *Managing innovation*. London: Sage.
- Fullan, M. (2002). The change. *Educational Leadership*, May, 16–20.
- Fullan, M. (2007). *The new meaning of educational change* (4e ed.). New York: Teachers College Press.
- Gannaway, D., Hinton, T., Berry, B., & Moore, K. (2013). Cultivating change: Disseminating innovation in higher education teaching and learning. *Innovations in Education and Teaching International*, 50(4), 410–421.
- George, L., & Sabapathy, T. (2011). Work motivation of teachers: Relationship with organizational commitment. *Canadian Social Science*, 7(1), 90.

- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Englewood Cliffs: Prentice Hall.
- Hasanefendic, S., Birkholz, J. M., Horta, H., & van der Sijde, P. (2017). Individuals in action: bringing about innovation in higher education. *European Journal of Higher Education*, 7(2), 101–119.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Janssen, O. (2000). Job demands, perceptions of effort-reward fairness and innovative work behaviour. *Journal of Occupational and Organizational Psychology*, 73, 287–302.
- Janssen, O. (2005). The joint impact of perceived influence and supervisor supportiveness on employee innovative behaviour. *Journal of Occupational and Organizational Psychology*, 78(4), 573–579.
- Janssen, O., Schoonebeek, G., & Looy, V. B. (1997). Cognitions of empowerment: the link between participative management and employees' innovative behaviour. *Gedrag en Organisatie*, 10(4), 175–194.
- Johnson, K., Hays, C., Center, H., & Daley, C. (2004). Building capacity and sustainable prevention innovations: a sustainability planning model. *Evaluation and Program Planning*, 27, 135–149.
- Kanter, R. (1988). When a thousand flowers bloom: structural, collective, and social conditions for innovation in organizations. *Research in Organizational Behaviour*, 10, 169–211.
- Kleysen, R. F., & Street, C. T. (2001). Toward a multi-dimensional measure of individual innovative behaviour. *Journal of Intellectual Capital*, 2(3), 284–296.
- Koeslag-Kreunen, M. G., Van der Klink, M. R., Van den Bossche, P., & Gijssels, W. H. (2018). Leadership for team learning: The case of university teacher teams. *Higher Education*, 75(2), 191–207.
- Krause, D. E. (2004). Influence-based leadership as a determinant of the inclination to innovate and of innovation-related behaviours: An empirical investigation. *The Leadership Quarterly*, 15(1), 79–102.
- Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? *Rasch Measurement Transactions*, 12, 636.
- Linacre, J. M. (2006). *Content Validity, Construct Validity and Misfitting Items*. Retrieved January 13th, 2006, from www.rasch.org/rmt/rmt121f.htm
- Linacre, J. M. (2017). Winsteps® (version 4.0.0). [computer software]. Beaverton, Oregon: [Winsteps.com](http://www.winsteps.com). Retrieved June, 2017. Available from <http://www.winsteps.com/>.
- Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, 8(2), 350.
- Loh, L. C., Friedman, S. R., & Burdick, W. P. (2013). Factors promoting sustainability of education innovations: A comparison of faculty perceptions and existing frameworks. *Education of Health*, 26(1), 32–38.
- Lomax, R. G., & Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling*. Psychologypress.
- Ludlow, L. H., Enterline, S. E., & Cochran-Smith, M. (2008). Learning to teach for social justice beliefs scale: An application of Rasch measurement principles. *Measurement and Evaluation in Counseling and Development*, 40(4), 194–214.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Messmann, G. M. A. (2012). *Innovative work behaviour: Investigating the nature and facilitation of vocational teachers' contributions to innovation development*. Regensburg: Universität Regensburg.
- Messmann, G., & Mulder, R. H. (2011). Innovative work behaviour in vocational colleges: Understanding how and why innovations are developed. *Vocations and Learning*, 4(1), 63–84.
- Messmann, G., & Mulder, R. H. (2012). Development of a measurement instrument for innovative work behaviour as a dynamic and context-bound construct. *Human Resource Development International*, 15(1), 43–59.
- Messmann, G., & Mulder, R. H. (2014). Exploring the role of target specificity in the facilitation of vocational teachers' innovative work behaviour. *Journal of Occupational and Organizational Psychology*, 87(1), 80–101.
- OECD (2009). *Creating effective teaching and learning environments: First results from TALIS, Teaching and Learning International Survey*. OECD publishing.
- OECD. (2018). *The future of education and skills: Education 2030*. OECD Education 2030.
- Pieterse, A. N., Van Knippenberg, D., Schippers, M., & Stam, D. (2010). Transformational and transactional leadership and innovative behaviour: The moderating role of psychological empowerment. *Journal of Organizational Behaviour*, 31(4), 609–623.
- Radaelli, G., Lettieri, E., Mura, M., & Spiller, N. (2014). Knowledge sharing and innovative work behaviour in healthcare: A micro-level investigation of direct and indirect effects. *Creativity and Innovation Management*, 23(4), 400–414.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Reay, T., Chreim, S., Golden-Biddle, K., Goodrick, E., Williams, B. E., Casebeer, A., & Hinings, C. R. (2013). Transforming new ideas into practice: An activity based perspective on the institutionalization of practices. *Journal of Management Studies*, 50(6), 963–990.

- 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.
- Schleicher, A. (2012). *Preparing Teachers and Developing School Leaders for the 21st Century: Lessons from around the World*. OECD publishing, 2, rue Andre Pascal, F-75775 Paris Cedex 16, France.
- Scott, S. G., & Bruce, R. A. (1994). Determinants of innovative behaviour: A path model of individual innovation in the workplace. *Academy of Management Journal*, 37(3), 580–607.
- Scott, S. G., & Bruce, R. A. (1998). Following the leader in R&D: The joint effect of subordinate problem-solving style and leader-member relations on innovative behaviour. *IEEE Transactions on Engineering Management*, 45(1), 3–10.
- Sherry, L. (2002). Sustainability of innovations. *Journal of Interactive Learning Research*, 13(3), 211.
- Sias, C. M., Nadelson, L. S., Juth, S. M., & Seifert, A. L. (2017). The best laid plans: Educational innovation in elementary teacher generated integrated STEM lesson plans. *The Journal of Educational Research*, 110(3), 227–238.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction* (no. 82). Sage.
- Spreitzer, G. M. (1995). Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of Management Journal*, 38(5), 1442–1465.
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston: Pearson.
- Thurlings, M., Evers, A. T., & Vermeulen, M. (2015). Toward a model of explaining teachers' innovative behaviour a literature review. *Review of Educational Research*, 85(3), 430–471.
- Trochim, W. M. K. (2002). Measurement validity types. Retrieved January 13th, 2006, from <http://www.socialresearchmethods.net/kb/measval.htm>
- Tuominen, T., & Toivonen, M. (2011). Studying innovation and change activities in KIBS through the lens of innovative behaviour. *International Journal of Innovation Management*, 15(02), 393–422.
- Van de Ven, A. H. (1986). Central problems in the management of innovation. *Management Science*, 32(5), 590–607.
- Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research*, 35(5), 441–461.
- West, M. A., & Farr, J. L. (1989). Innovation at work: psychological perspectives. *Social behaviour*, 4, 15–30.
- Wolbers, K. A., Dostal, H. M., Skerrit, P., & Stephenson, B. (2017). The impact of three years of professional development on knowledge and implementation. *The Journal of Educational Research*, 110(1), 61–71.
- Wolfe, E. W., & Smith, J. E. (2007). Instrument development tools and activities for measure validation using Rasch models: part II—validation activities. *Journal of applied measurement*, 8(2), 204–234.
- Woodman, R. W., Sawyer, J. E., & Griffin, R. W. (1993). Toward a theory of Organizational Creativity. *Academy of management review*, 18(2), 293–321.
- World Economic Forum. (2016). *The future of jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*. New York: World Economic Forum LLC.
- Yidong, T., & Xinxin, L. (2013). How ethical leadership influence employees' innovative work behaviour: A perspective of intrinsic motivation. *Journal of Business Ethics*, 116(2), 441–455.

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

MSc. Peggy Lambriex-Schmitz is a PhD student at the department of Educational Research and Development of the University Maastricht School of Business and Economics (the Netherlands), and a researcher at Research Centre Educational Innovation and Continuous Professional Development, Zuyd University of Applied Sciences, Heerlen (the Netherlands). Her main research interests concern implementing sustainable innovations in higher and secondary vocational education, curriculum development, fostering innovative work behaviour and professionalization of teachers. She is a senior lecturer, educational scientist and consultant at Zuyd University.

Dr. Marcel van der Klink is Professor at Research Centre of Educational Innovation and Continuous Professional Development at the Zuyd University of Applied Sciences Heerlen (The Netherlands). His research focuses on learning and working of teachers in vocational and higher education, curriculum design, student retention and the meaning of twenty-first century skills for higher education.

Dr. Simon Beusaert is Associate Professor in the Field of workplace learning at Maastricht University, (the Netherlands). His current research focuses on supporting (in)formal learning and assessment for learning at the workplace, and their relation with employees' professional development, employability and innovative working behaviour.

Dr. Monique Bijker is Senior Researcher at the Research Centre Educational Innovation and Continuous Professional Development at the Zuyd University of Applied Sciences Heerlen (The Netherlands). She focuses on the development and validation of psychological measurement instruments and assessments, using techniques such as CFA, SEM, and Rasch modeling.

Prof. Dr. Mien Segers is Professor of Corporate Learning at the department of Educational Research and Development of the University Maastricht School of Business and Economics (the Netherlands). Her major research interests are workplace learning, formal as well as informal learning of professionals, and its role as a lever for innovation.

Affiliations

Peggy Lambriex-Schmitz¹ · Marcel R. Van der Klink^{1,2} · Simon Beusaert² · Monique Bijker¹ · Mien Segers²

Marcel R. Van der Klink
marcel.vanderklink@zuyd.nl

Simon Beusaert
s.beusaert@maastrichtuniversity.nl

Monique Bijker
Monique.bijker@zuyd.nl

Mien Segers
m.segers@maastrichtuniversity.nl

¹ Research Centre Educational Innovation and Continuous Professional Development, Zuyd University of Applied Sciences, P.O. Box 550, 6400 AN Heerlen, The Netherlands

² Department of Educational Research and Development, School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands