Teaching and Teacher Education 45 (2015) 128–136

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Primary teachers' attitudes, intentions and practices regarding formative assessment

Zi Yan[*, 1], Eric Chi Keung Cheng [2]

Department of Curriculum and Instruction, The Hong Kong Institute of Education, Hong Kong

HIGHLIGHTS

- The TPB model was used to explain teachers' FA intentions and practices.
- Instrumental attitude, subjective norm, and self-efficacy predict FA intentions.
- Affective attitude and controllability do not predict FA intentions.
- The TPB components are not effective predictors of FA practices.
- Contextual factors should be considered when examining FA practices.

ARTICLE INFO

Article history:
Received 21 April 2014
Received in revised form
30 September 2014
Accepted 6 October 2014
Available online

Keywords:
Formative assessment
Teacher
Theory of planned behaviour
Rasch measurement
Path analysis

ABSTRACT

This study aims to explore the relationships among teachers' attitudes, intentions, and practices regarding formative assessment under the framework of the Theory of Planned Behaviour. A total of 450 teachers from 10 primary schools were surveyed. Teachers' responses to the scales were calibrated using Rasch analysis and then subjected to path analysis. Instrumental attitude, subjective norm, and self-efficacy were significant predictors of teachers' intentions to conduct formative assessment. However, the TPB components were not effective predictors of teachers' formative assessment practices. Teachers' awareness of formative assessment was articulated and implications for teacher training as well as further studies were discussed.

1. Introduction

It is a well-received notion that school assessments serve two major purposes: summative and formative. Given that both purposes have their own irreplaceable functions in education, a dilemma faced by many teachers is how to reconcile the role of summative and formative assessments in teaching. In many examination-oriented education systems, an often-heard claim is that the pressure in schools to improve students' achievement in high-stakes summative examinations precludes the use of formative assessment (Wiliam, Lee, Harrison, & Black, 2004).

Given that the first and foremost purpose of assessment in education is to support learning (Black & Wiliam, 2006), the high status of formative assessment is reflected in recent documents regarding Hong Kong's educational reform. The Curriculum Development Council (CDC, 2001) has urged a change in assessment practices to a system in which schools place more emphasis on formative assessment, making the latter an integral part of classroom teaching. CDC and the Hong Kong Examinations and Assessment Authority (CDC & HKEAA, 2007) also recommended that formative assessment should be given more attention and have a higher status than summative assessment due to its compelling educative functions. However, this top-down reform initiative relating to assessment is vulnerable as it ignores the role of teachers (Rink & Mitchell, 2002) and deficient teacher cooperation, knowledge, or beliefs (Brown, Hui, Yu, & Kennedy, 2011). It is

* Corresponding author. D1-1/F-49, Department of Curriculum and Instruction, The Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, N.T., Hong Kong. Tel.: +852 2948 6367; fax: +852 2948 7563.
E-mail address: zyan@ied.edu.hk (Z. Yan).
[1] Assistant Professor at the Department of Curriculum and Instruction at the Hong Kong Institute of Education. His main research interests are in Rasch measurement, assessment, and scale development.
[2] Associate Professor at the Department of Curriculum and instruction at the Hong Kong Institute of Education. His main research interests are in knowledge management, learning study and school management.

http://dx.doi.org/10.1016/j.tate.2014.10.002
0742-051X/

evident that successful educational reform relies on teachers' understanding, participation, and support for change (Hallinger, 2011) as teachers' views have a significant impact on the implementation of educational reform. However, few attempts have been made to build a structural understanding of teachers' attitudes, intentions, and practices regarding formative assessment. In other words, there is limited knowledge about the factors contributing to the formation of teachers' intentions to conduct formative assessment as well as the influences of attitudinal factors on teachers' formative assessment practices. The present study aims to fill this gap by exploring the relationships among teachers' attitudes, intentions, and practices regarding formative assessment under the framework of the Theory of Planned Behaviour (TPB).

1.1. Literature review

Summative assessment is the practice of collecting information with a view to summarising how much learning has taken place. On the other hand, formative assessment is defined by Black and Wiliam (1998a) as "encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 7). It emphasises the learning process and is usually conducted on a daily basis.

Formative assessment is well believed to be productive in optimising teaching practice in ways that support student learning. Plenty of evidence can be found in literature that formative assessment practices can result in improved student achievement and reduce the achievement gap among students. For example, in Hattie's (2009, 2012) research on "visible learning", formative assessment was rated as one of the most effective methods, having a visible effect on student achievement. Wiliam et al. (2004) also found that integration of formative assessment into teaching results in enhanced student learning performance. In a review of empirical studies concerned with classroom assessment, Black and Wiliam (1998a) reported consistent learning gains with surprisingly large effect sizes for students when assessment practices were well designed and used in a formative approach.

The major principles of formative assessment – identifying students' weaknesses and strengths, enhancing students' motivation and metacognition, providing feedback to inform teaching and learning – are well recognised in their capacity for improving students' learning (Black & Wiliam, 1998a; Wiliam et al., 2004). Both teachers and students can benefit from formative assessment by acquiring learning data which can be used to support personalised learning.

Although formative assessment is well accepted with regard to enhancing teaching and learning, it is a demanding task for teachers due to the uncertainty and flexibility of formative assessment (Bell & Cowie, 2001). Black and Wiliam (1998b) also stated that formative assessment is by no means a simple matter and "there is no 'quick fix' that can be added to existing practice with promise of rapid reward" (p. 15). Teaching is a highly personal activity where teachers enact their educational philosophies and make sense of notions of curriculum and assessment (Harrison, 2013). Therefore, attitudinal and cognitive factors may have important impacts on teachers' teaching practices in the classroom.

Researchers have done impressive studies assessing teachers' attitudes and practices regarding assessment. However, most studies have focused on only one aspect of assessment (i.e., attitudes or practices) (e.g., Brown, 2004; Brown et al., 2011; Dixon & Haigh, 2009; Wong, 2014). Other research has studied both teachers' attitudes and practices regarding assessment but failed to investigate the correlation between these variables under a sound theoretical framework (e.g., Brown, Kennedy, Fok, Chan, & Yu, 2009; Büyükkarcı, 2014; Duwairi, 2013).

1.2. Theory of planned behaviour

The TPB proposed by Ajzen (1985, 1991) is a rigorous theoretical framework which has the potential to provide prediction and explanation of teachers' intentions to and practice of formative assessment. TPB has been applied successfully in attempts to provide a better interpretation of diverse behaviours in western settings (Armitage & Conner, 2001; Kersaint, Lewis, Potter, & Meisels, 2007; Lipnevich, MacCann, Krumm, Burrus, & Roberts, 2011; MacFarlane & Woolfson, 2013) and the Hong Kong context (Dodgson, Henly, Duckett, & Tarrant, 2003; Mok & Lee, 2013; Yan & Sin, 2014). TPB outlines relationships among attitude, subjective norms, perceived behavioural control, intention, and practice of a particular behaviour. According to TPB theory, three determinants – including attitude towards the behaviour, subjective norms, and perceived behavioural control – exert their effects on behaviour through intentions (Ajzen, 1991).Generally, those with a favourable attitude, positive subjective norms, and high level of perceived behavioural control will more likely have the intention to perform the behaviour. The level of intention and the degree of perceived behavioural control will determine the individual's actual behaviour. The theoretical model of TPB is presented in Fig. 1.

Attitude towards the behaviour is defined as the personal orientation towards performing the behaviour and that orientation often contains two components: affective and instrumental attitude (Ajzen, 1991, 2002a). The affective attitude incorporates feelings or emotions arising from the prospect of performing the behaviour; while instrumental attitude carries an appraisal of the consequences of performing the behaviour. Subjective norm refers to perceived social pressure from significant others to perform (or not perform) the behaviour. Perceived behavioural control refers to one's perception of the difficulty of and control over performing the behaviour. Self-efficacy is a widely used indicator of perceived difficulty. Ajzen (2002a) suggested that a behaviour's controllability, which refers to people's beliefs that they have control over the behaviour, is another important indicator of perceived behavioural control. A meta-analysis on 11 empirical studies applying TPB also supported the distinction between self-efficacy and controllability across a broad range of behaviours (Trafimow, Sheeran, Conner, & Finlay, 2002).

1.3. Aims of the study

This study aims at examining the extent to which TPB can predict and explain teachers' intentions to and practices of formative assessment in the classroom. The following two specific hypotheses are tested:

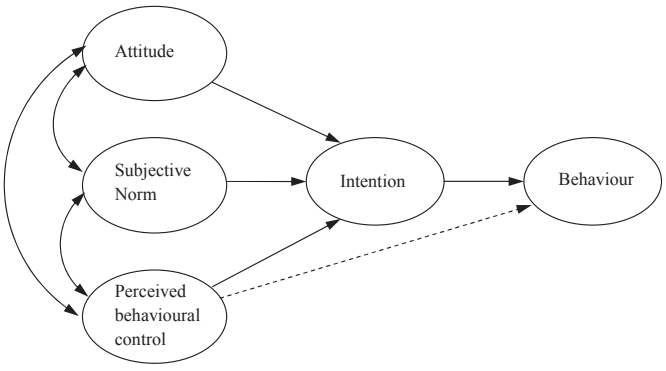

Fig. 1. The model of theory of planned behaviour (Ajzen, 1991).

(1) Teachers' intentions to formative assessment can be predicted by attitude, subjective norm, and perceived behavioural control regarding formative assessment; and
(2) Teachers' formative assessment practices can be predicted by intentions and perceived behavioural control regarding formative assessment.

2. Methodology and methods

2.1. Participants

The data was collected from 10 Hong Kong primary schools that participated in a Quality Education Fund project supported by the Hong Kong government. Although a convenience sample was used, the 10 participating schools were carefully selected to represent different school bandings and funding modes. A total of 535 teachers from participating schools were invited to participate in the survey and 450 (84.1%) completed questionnaires were returned. The teacher sample consisted of 95 (21.1%) males and 349 (77.6%) females, with 6 (1.3%) providing no gender information. With regard to teaching experience, 189 (42.0%) teachers had no more than 10 years of experience; 187 (41.6%) had 11–20 years of experience; 48 (10.7%) were experienced teachers with more than 20 years of experience; 26 (5.8%) teachers did not provide such information.

2.2. Instruments

A Teacher's Conceptions and Practices of Formative Assessment Questionnaire containing seven scales was developed to assess the five components in the TPB framework regarding formative assessment. The scales included the affective attitude scale (AAT), instrumental attitude scale (IAT), subjective norm scale (SNO), controllability scale (CON), self-efficacy scale (SEF), intention scale (INT), and behaviour scale (BEH). The scale development procedure was guided by DeVellis' (2012) instruction and item crafting was guided by Ajzen's (2002a) principles for TPB scale construction. The item pool was constructed based on previous literature on teachers' conceptions, beliefs, and practices regarding assessment (e.g., Brown et al., 2011; Yan, 2014), and consultative discussions with relevant experts (in the field of educational assessment) and front-line teachers who have had first-hand experiences of formative assessment in classroom teaching. The scales were then subjected to a pilot test on a small sample of students ($N = 21$). Admittedly, it is difficult to conduct meaningful statistical analysis on such a small sample. However, participants' responses to items as well as their written comments helped refine the scales by identifying ambiguities and anomalies in item wording, as well as possible bias.

The AAT scale, consisting of 7 items (e.g., I like Formative Assessment), examined teachers' feelings or emotions aroused by conducting formative assessment. The IAT scale, with 13 items (e.g., Formative Assessment can raise students' interest in learning), was used to tap teachers' appraisal of how advantageous the consequences of conducting formative assessment would be. Five items (e.g., As far as I know, the principal of my school believes that formative assessment should be implemented) were included in the SNO scale to assess teachers' perceptions of important others' opinions regarding formative assessment. Teachers' beliefs about their control over conducting formative assessment were investigated by the CON scale, which comprised 4 items (e.g., I can decide whether or not to implement formative assessment). The SEF scale, tapping teachers' self-efficacy in conducting formative assessment, had 6 items (e.g., I can design appropriate assessment tasks for formative assessment). Teachers' intentions to formative assessment were assessed by a 6-item INT scale (e.g., I am willing to integrate formative assessment into my teaching). Teachers' formative assessment practices were gauged by two items in the BEH scale (e.g., In the past six months, how often have you implemented formative assessment?).

An operational definition of formative assessment was printed at the head of the questionnaire in order to provide a common basis for teachers' interpretation. This definition referred to formative assessment as being assessment activities undertaken by teachers and/or by their students in the daily teaching and learning process. These assessment activities provide information to be used as feedback to modify teaching and learning activities. Participants were asked to rate each item on a 6-point scale with response options ranging from Strongly Disagree (1) to Strongly Agree (6). The scores on negatively-worded items were reversed before the data analysis so as to maintain the consistency of interpretation, i.e., a higher score represents a higher level on the latent trait under investigation.

2.3. Data analysis

Two analytical methods, i.e., Rasch analysis (Rasch, 1960) and path analysis, were used in the present study. Rasch rating scale analysis using Winsteps 3.7 (Linacre, 2011) was first used for examining the psychometric properties of the seven scales and for calibrating teachers' (person) measures on each of the seven latent traits, i.e., ATT, IAT, SNO, CON, SEF, INT, and BEH. The Rasch-calibrated person measures were subsequently analysed by path analysis using Amos 20 (Arbuckle, 2011) to investigate the relationships among the TPB components. This approach to data analysis differs from the conventional structural equation model in that it applies Rasch analysis for measuring the latent (unobserved) traits, whereas the conventional structural equation model usually uses confirmatory factor analysis to fulfil the function of measurement model.

An inherent weakness associated with conventional analytical techniques based on classic test theory, such as factor analysis, is that they require linear, interval scale data input (Wright, 1997). Raw data collected through Likert-type scales, however, are always ordinal since the categories of Likert-type scales indicate only ordering without any proportional levels of meaning (Bond & Fox, 2007; Wright, 1997). Therefore, it is highly possible to arrive at misleading conclusions if applying conventional analytical techniques to raw scores which are ordinal data in nature. The Rasch model overcomes this problem by converting ordinal data into interval measures which have a constant interval meaning and, therefore, provides objective measurement from ordered category responses (Linacre, 2006). Once the interval metric is established, person measures and item difficulties are to be calibrated onto a single unidimensional latent trait scale which facilitates direct comparisons between person measures and item difficulties. Furthermore, in Rasch analysis, person measures are independent from the items used, with item difficulties being independent from the sample recruited because the estimates are calibrated within a common metric rather than against a single test situation (for person ability estimates) or a particular sample of test takers (for item difficulty estimates). Empirically, Rasch analysis has been successfully applied in education and social sciences in addressing assessment issues (Bond & Fox, 2007; Panayides, Robinson, & Tymms, 2010; Törmäkangas, 2011).

Multiple criteria – including Rasch person/item reliability, item fit statistics, the amount of variance explained by each of the scale measures, and step threshold – are used to examine the psychometric properties of the seven scales. Rasch person/item reliability estimates the replicability of person/item ordering along the latent trait metric (Bond & Fox, 2007). Item fit statistics estimate the

extent to which the data matches the measurement specifications of the Rasch model. Outfit and Infit mean squares (MNSQ) are widely used indices of item fit statistics. The values of Outfit and Infit MNSQ range from 0 to positive infinity with 1.0 indicating the (unattainable) perfect fit to the Rasch model. Researchers (for example, Wright & Linacre, 1994) suggested that MNSQs falling in the range of 0.6–1.4 indicated a productive measurement for survey data with rating scales. This criterion was adopted as the cut-off value of MNSQ fit statistics in this study. Variance explained by Rasch measures refers to the proportion of variance in the observed data which can be explained by the Rasch measures (Linacre, 2006). A higher proportion of variance indicates that the Rasch model better predicts both items and persons. Step thresholds are examined to ensure the appropriate category functioning of the rating scales.

In path analysis, multiple indices, in addition to traditional chi-square (χ^2), are used to examine the goodness of model-data fit including goodness of fit (GFI), adjusted goodness of fit (AGFI), comparative fit index (CFI), the Tucker-Lewis-Index (TLI), and the standardised root-mean-square error of approximation (RMSEA). The value of GFI, AGFI, CFI, and TLI can vary from 0 to 1, and values greater than 0.90 (Garson, 2009) indicate a good fit. The RMSEA index takes into account sample size when evaluating the fit, with low values indicating good fit. As a general rule, RMSEA less than 0.06 (Hu & Bentler, 1998, 1999) can be considered to be an indication of a good model fit.

3. Results

3.1. Psychometric properties of the scales

The psychometric properties of the seven scales were first investigated from a Rasch measurement perspective. The Rasch analysis identified item 3 (Formative assessment increases academic burdens on students) and item 9 (Formative assessment increases my workload) from the IAT scale as misfitting to the Rasch model (both Infit and Outfit MNSQ being higher than 1.4). According to Hart and Wright (2002), deleting items from the scale may affect the accuracy of estimations of the remaining items and the error rate of model estimates. Therefore, the identified misfitting items were removed from the scale one at a time according to the misfit order, with Rasch analysis re-applied until all remaining items showed sufficient fit to the Rasch model. After removing items 3 and 9 from the IAT scale, one more item (item 1: Formative assessment can offer a comprehensive appraisal of students' performance) appeared misfitting to the model. This item was subsequently removed from the scale which resulted in a 10-item IAT scale.

One item of the SNO scale (item 1: As far as I know, officials of the Education Bureau believe that formative assessment should be implemented) showed marginal fit to the Rasch model (Infit MNSQ = 1.42). Given that Outfit MNSQ of this item was lower than 1.4 and consideration of officials of the Education Bureau's opinions is an important factor in influencing teachers' teaching, this item was kept in the SNO scale. All items in the other five scales showed sufficient fit to the Rasch model. Table 1 presents the summary of psychometric properties of all scales and the final version of the scales are attached in Appendix 1.

It can be seen from Table 1 that the Rasch person/item reliabilities for all scales are higher than 0.80 except the person reliability for the SNO scale. Rasch measures explained over 50% of the variables observed in the data for all scales except the SNO scale, in which Rasch measures explained 47.3% of the variance. The results indicated that most of the scales had quite good psychometric properties while the SNO scale only showed an acceptable quality. The category functioning of the seven rating scales were examined to determine whether respondents used all response opportunities appropriately. It can be seen that the step thresholds (the intersection point between consecutive categories) advanced monotonically with the category, indicating that the 6-point rating scale functioned well, i.e., higher performance categories corresponded to higher measures of the latent trait. In summary, the results showed that the scales were psychometrically robust enough for use with the sample in the present study.

Table 1
Psychometric properties of measurement scales.

Scale	No. of items	Rasch person/Item reliability	Variance explained by measures	Step threshold				
				Step 1	Step 2	Step 3	Step 4	Step 5
AAT	7	0.87/0.99	60.2%	−4.20	−3.58	−0.92	2.18	6.52
IAT	10	0.88/0.98	51.6%	−3.63	−3.02	−1.09	1.87	5.86
SNO	5	0.75/0.98	47.3%	−2.38	−1.44	−0.91	0.83	3.90
CON	4	0.85/0.98	70.8%	−4.38	−3.18	−1.26	1.46	7.31
SEF	6	0.84/0.96	60.6%	−3.81	−2.85	−0.97	1.73	5.91
INT	6	0.88/0.93	80.9%	−10.38	−6.68	−2.40	5.60	13.86
BEH	2	0.82/0.99	81.1%	−11.88	−3.74	0.21	4.78	10.63

3.2. Descriptive satistics

Descriptive statistics were undertaken to provide a general picture of the interval Rasch-calibrated measures of teachers on the seven constructs. Table 2 presents means and standard deviations of teachers' measures on the scales as well as Pearson correlations among the constructs of interest.

In Rasch analysis, the mean of item difficulties is arbitrarily set to zero and the interpretation of item difficulties and person measures are based on pair-wise comparisons between items and persons. Therefore, person measures higher than zero indicate a positive response, while person measures lower than zero indicate a negative response. It can be seen from Table 2 that, in general, teachers held a substantially high level of intention to conduct formative assessment (INT Mean = 6.01). Attitude, both affective (AAT Mean = 1.02) and instrumental (IAT Mean = 1.79), and subject norm (SNO Mean = 1.40) regarding formative assessment were quite positive; while perceived behavioural control, i.e., controllability (CON Mean = 0.06) and self-efficacy (SEF Mean = 0.54), were just slightly positive. It is worth noting that teachers had a slightly negative mean measure on behaviour (BEH Mean = −0.79). This indicates that, according to teachers' perceptions, formative assessment was not frequently conducted in their classrooms. All variables except BEH showed median to high, and statistically significant, correlations among one another. BEH had relatively weak correlations with other variables.

Table 2
Means, standard deviations, and correlations of TPB components.

Component	M	SD	AAT	IAT	SNO	CON	SEF	INT
AAT	1.02	2.22	–					
IAT	1.79	1.91	0.82**	–				
SNO	1.40	1.70	0.41**	0.47**	–			
CON	0.06	2.86	0.53**	0.53**	0.28**	–		
SEF	0.54	2.07	0.53**	0.56**	0.32**	0.74**	–	
INT	6.01	5.30	0.50**	0.60**	0.41**	0.51**	0.64**	–
BEH	−0.79	4.78	0.07	0.12*	0.17**	0.15**	0.20**	0.23**

Note. *$p < .05$; **$p < .01$.

3.3. Results of path analysis

Teachers' Rasch-calibrated measures on the TPB components are subsequently subjected to path analysis, aiming at addressing the main research questions – to explore whether teachers' intentions to and practices of formative assessment can be predicted by attitude, subjective norm, and perceived behavioural control regarding formative assessment. According to Allison's (2003) suggestion, the maximum likelihood methods with the expectation maximisation algorithm (Dempster, Laird, & Rubin, 1977) was used to handle missing data. The parameter estimates generated by maximum likelihood methods excelled those from other conventional methods (e.g., listwise deletion and pairwise deletion) in terms of consistency, asymptotic efficiency, and asymptotic normality (Allison, 2003). The path analysis was then applied to the complete set of data. The results of path analysis showed a satisfactory fit between the proposed TPB-based model and the observed data. The chi-square statistic ($\chi^2 = 7.678$, df = 3, $p = .053$) and the relative chi-square ($\chi^2/df = 2.559$) were satisfactory. The other fit statistics (GFI = 0.995, AGFI = 0.953, CFI = 0.997, TLI = 0.971) were all higher than 0.95. RMSEA (0.059) was less than 0.6, also indicating a fairly good fit. These results supported the appropriateness of the proposed TPB-based model in predicting Hong Kong teachers' intentions and practices regarding formative assessment.

Standardised regression weights of the paths from IAT, SNO, and SEF to INT were significant ($p < .01$). The strongest predictor of INT was SEF ($\beta = 0.44$), followed by IAT ($\beta = 0.34$) and SNO ($\beta = 0.14$). The paths from AAT ($\beta = -0.06$) and CON ($\beta = -0.01$) to INT were not significant and the correlation between CON and INT was virtually zero. The standardised regression weight of the path from INT to BEH ($\beta = 0.14$) was significant ($p < .01$), while the paths from perceived behavioural control components, i.e., CON ($\beta = -0.01$) and SEF ($\beta = -0.09$), were not significant and the correlation between CON and BEH was virtually zero. The proposed model accounts for 51% of the variances in teachers' intentions to formative assessment, but only accounts for 6% of the variances in teachers' practices of formative assessment. The relationships among the latent traits in the TPB-based model are presented in Fig. 2.

The direct and indirect effect coefficients of predictors on formative assessment practices (BEH) are presented in Table 3. In the proposed model, INT (via direct effect) and SEF (via direct and indirect) had a similar effect on BEH, while the other predictors had

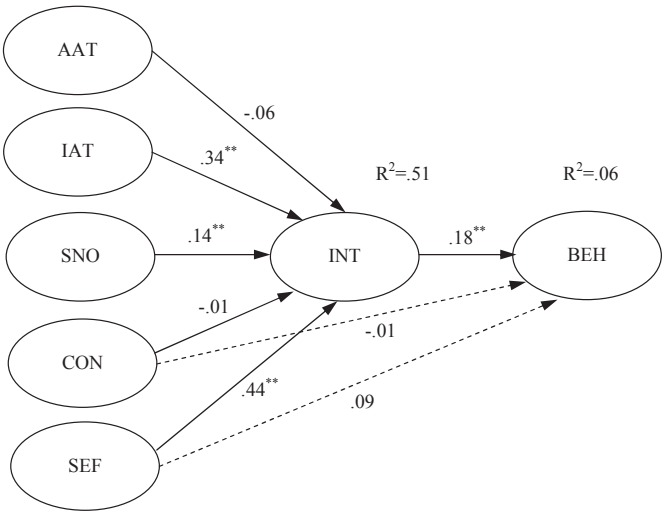

Fig. 2. Path analysis based on the theory of planned behaviour (**$p < .01$).

Table 3
Effects of each TPB component on teachers' formative assessment practices.

Variables	Direct effect	Indirect effect	Total effect
AAT	–	−0.01	−0.01
IAT	–	0.06	0.06
SNO	–	0.03	0.03
CON	−0.01	−0.00	−0.01
SEF	0.10	0.08	0.18
INT	0.18	–	0.18

a far less than substantial effect on BEH. The results indicated that TPB components had weak effects on teachers' formative assessment practices.

4. Discussion

This study used TPB as a theoretical framework to examine whether (1) teachers' intentions to formative assessment can be predicted by attitude, subjective norm, and perceived behavioural control regarding formative assessment; and (2) teachers' formative assessment practices can be predicted by intention and perceived behavioural control regarding formative assessment. Much past research has investigated the benefits that formative assessment may bring and the way in which formative assessment should be conducted, but few attempts have been made to build a structural understanding of formative assessment or of the relationships among variables which have influences on teachers' intentions and practices regarding formative assessment. The findings of this study showed that the first hypothesis was generally supported. Teachers' intentions to formative assessment can be predicted by attitude, subjective norm, and perceived behavioural control regarding formative assessment. The TPB components accounted for around 51% of the variances in teachers' intentions to formative assessment. However, the second hypothesis was not supported by the results. Teachers' intention and perceived behavioural control regarding formative assessment only accounted for 6% of the variances in teachers' formative assessment practices.

The results showed that instrumental attitude, subjective norm, and self-efficacy were significant predictors of teachers' intentions to conduct formative assessment. Teachers who had a favourable instrumental attitude, a positive subjective norm, and a high level of self-efficacy were more likely to have the intention to conduct formative assessment. This result supported the appropriateness of TPB as a theoretical framework in explaining the formation of teachers' intentions to formative assessment.

Teachers' intentions were most strongly predicted by self-efficacy ($\beta = 0.44$); and strongly predicted by instrumental attitude ($\beta = 0.34$). The predictive power of subjective norm ($\beta = 0.14$) on intention was relatively weak, although still significant. In Armitage and Conner's (2001) meta-analysis on 185 studies that applied TPB, attitudes were the strongest predictor of intentions ($\beta = 0.49$), followed by perceived behavioural control ($\beta = 0.43$) and subjective norm ($\beta = 0.34$). This study is consistent with past research in that the correlation between subjective norm (social pressure or important others' opinions) and intention is the weakest. It implies that teachers' intentions to conduct formative assessment are determined by internal factors more than external factors, such as important others' opinions. However, this study differs from Armitage and Conner's (2001) report in that self-efficacy showed a stronger impact on intention than instrumental attitude. This finding is similar to Yan's (2014) study which focused on another assessment issue (school-based assessment) in the Hong Kong context. Yan (2014) reported that self-efficacy had a

stronger predicting power on teachers' intentions to conduct school-based assessment than did instrumental attitude. Dixon and Haigh (2009) also observed that teachers with higher self-efficacy were more willing to try new initiatives in assessment. It would appear that when considering assessment practices, teachers regard their own capacities, i.e., whether they have the necessary skills, as the first concern. Teachers are more likely to conduct the assessment practices when and if they feel confident with that particular assessment approach.

The results revealed that affective attitude and controllability were not significant predictors of teachers' intentions to conduct formative assessment. The fact that affective attitude had weak correlation with teachers' intention is not a surprise according to Yan's (2014) explanations. For those studies (e.g., Courneya, Vallance, Jones, & Reiman, 2005; Lawton, Conner, & Parker, 2007; Rhodes & Courneya, 2003; Trafimow et al., 2004) which found affective attitude to be strongly influential on intention, the behaviours under investigation were usually personal behaviours (e.g., exercise, smoking, and speeding while driving) which generally had personal consequences. By contrast, working behaviours in the classroom, including formative assessment, have consequences on both teachers and their students. Therefore, the perceptions of a given behaviour's consequences on students' learning (instrumental component) rather than personal feelings or emotions (affective component) appear to be more important in determining teachers' intentions to formative assessment.

Unlike self-efficacy, the other perceived behavioural control component – controllability – was not a significant predictor of intention. This is consistent with the claim made by Trafimow et al. (2002) that self-efficacy was superior to controllability in predicting intention and behaviour. Although the reasoning is inevitably speculative, a possible explanation is that difficulty of the behaviour is a more important issue than control over the behaviour when teachers consider conducting formative assessment. Although, theoretically, both self-efficacy and controllability can reflect internal as well as external factors (Ajzen, 2002b), empirically the items in the self-efficacy scale were gauging internal beliefs on one's competence (e.g., I can integrate formative assessment into the teaching and learning process; I can design appropriate assessment tasks for formative assessment) while items in the controllability scale were more easily influenced by external factors (e.g., I can decide the frequency of implementing formative assessment; I can decide the timing of implementing formative assessment). The level of teachers' perceived control might be largely influenced by external factors, such as the timetable, the school culture, etc. Teachers who felt they had control were not necessarily more willing to conduct formative assessment if s/he was not confident in her or his competence. In contrast, teachers who possess less control might still be willing to (that is, only intention, not action) conduct formative assessment if s/he was confident. As a consequence, perceived control (i.e., controllability) appears less predictive than perceived difficulty (i.e., self-efficacy).

Although the TPB components can well predict teachers' intentions to formative assessment, the proposed TPB-based model did not explain teachers' reported formative assessment practices very well. According to Armitage and Conner's (2001) meta-analysis on studies that applied TPB, an average of 27% of the variance in behaviour could be attributed to intentions and perceptions of behavioural control. In contrast, this study found that teachers' intentions and perceived behavioural control only accounted for 6% of the variances in teachers' formative assessment practices.

Such finding evidenced that even though teachers had quite a high level of intention to formative assessment and positive perceived control, their self-report showed that they did not conduct formative assessment frequently. It appears that there are factors, other than the TPB components (i.e., intentions and perceptions of behavioural control), influencing teachers' formative assessment practices. This is a support to the claim that the important objective contextual factors, in addition to the subjective psychological determinants, should not be excluded when using TPB to explain behaviours (Davis, Ajzen, Saunders, & Williams, 2002), especially for investigations on working behaviour like formative assessment. Compared to personal behaviours over which one has large volitional control, working behaviours are more susceptible to external or contextual variables. Some factors could easily be identified as hurdles to teachers' implementation of formative assessment, such as large class size and heavy workload. If a teacher has the care of up to 30 students in a class and teaches for more than 20 h per week, the implementation of formative assessment can be very difficult. Furthermore, as Harrison and Howard (2009) argued, effective formative practice has to be developed in appropriate classroom climate and dialogue, which are complex factors and are not solely controlled by the teacher.

Another likely, and more fundamental, explanation may be found in the relationship between the formative and summative/accountability purposes of assessment in school contexts. As Black and Wiliam (1998a) pointed out, one of the challenges in conducting formative assessment in the classroom is the powerful influence exerted by the requirements for certification and accountability. Although educational bureaucracies are trying to develop assessment structures for high-stakes accountability based on the principles of formative assessment (Hay, 2006), the global education system still emphasises accountability and evidence and requires schools and teachers to be responsible for outcomes (Rink & Williams, 2003). The instructional effects of formative assessment are related to students' learning ability, which will eventually result in better learning performance. However, if teachers are forced to put more emphasis on external high-stakes examinations within a relatively short timeframe, the spoon-feeding method of teaching and drilling, rather than adaptive instruction with formative assessment, may appear to be more attractive. Teachers, therefore, are likely to give up the adoption of formative assessment due to the pressure of external high-stakes examinations even though they are well aware of the educational benefits of formative assessment.

4.1. Implications on teacher training

The findings have important implications on professional development for teachers. The Hong Kong government has invested a huge amount of resources with the aim of changing the assessment culture in classrooms. For this to happen, teachers have to change their own conceptions of assessment and form intentions to the change. The findings in this study imply that the formation of teachers' intentions to formative assessment requires positive instrumental attitude towards formative assessment, and enhanced teachers' self-efficacy in conducting formative assessment.

Positive instrumental attitude can be built on recognition of the advantages of formative assessment. If teachers are aware that formative assessment can help them identify students' learning difficulties and utilise the assessment results to redesign teaching to enhance students learning, they will be much more eager to conduct formative assessment throughout the student learning process. Self-efficacy can be enhanced by equipping teachers with necessary knowledge and skills. These knowledge and skills should be included in professional training programmes which aim to optimise teachers' teaching practices. Furthermore, teacher

training should provide teachers with mastery experiences in formative assessment practice as such experience plays a crucial role in the formation of self-efficacy (Bandura, 1997; Mulholland & Wallace, 2001). If the new initiatives in formative assessment are not immediately successful, continuous professional support to teachers is especially important so as not to diminish teachers' self-efficacy by experiences of failure.

This study echoed Ofsted's (2007) observation that effective formative assessment in the classroom has not yet been achieved at a large scale even though teachers are aware of the benefits of formative assessment. It seems that there is a worldwide concern about teachers' poor understanding and weak adoption of formative assessment in classroom practice (Black & Wiliam, 1998a; Tan & Towndrow, 2009). Why have teachers' positive attitudes and beliefs regarding formative assessment not been fully translated in their classroom practice? One of the reasons might be the misconception about the role of formative assessment in regular instruction. Teachers probably still regard formative assessment as an added component, which needs extra time and resource, rather than an integrated part of regular instruction. A suggestion for teacher training might be generated from this study: to change teachers' beliefs on the relationship between formative assessment and regular instruction. Teacher training should strive to foster such a change in teachers' minds, as well as to equip them with the necessary skills, so that teachers are willing and capable to treat formative assessment as an integrated part of regular instruction rather than an added component that competes with other components for teaching time. As advocated by Smith (2013), there is an urgent need for more research-based knowledge on using assessment to sharpen teaching which, in turn, will bring positive outcomes in summative assessments. The public, including teachers themselves, needs to be convinced by research-based evidence of the effectiveness of formative assessment in enhancing teaching and learning, both long-term and short-term. Furthermore, how to make meaningful connections between formative and summative assessment may be another important topic of teacher training. As Maclellan (2004) pointed out, teachers' knowledge about assessment was largely compartmentalised and the connection between formative-summative modes of assessment were not well developed. Teacher training might need to develop teachers' capacity in collecting assessment data for both purposes, or make good use of data collected in one mode to the other. Such capacity might facilitate transforming teachers' intentions to formative assessment into actual practice.

4.2. Limitations

The interpretation and generalisation of the conclusions of this study should be carried out with caution. First, the data of this study were collected in the context of a single jurisdiction – Hong Kong. It is worth remembering that any attempt to generalise the findings of this study, especially in the prediction of formative assessment practices, should be made in education systems with similar social and cultural contexts. Second, a convenience sample was used in this study. Although the participating schools were carefully selected to represent different school bandings and funding modes, the generalisation is not as strong as that from a random sample. Third, teachers' formative assessment practices were investigated by self-report data. Future studies might consider collecting objective measures of teachers' practices that are free from the response-biases associated with self-report measures.

5. Conclusions

This study most significantly contributes to building a structural understanding of teachers' attitudes, intentions, and practices regarding formative assessment. This study supported that TPB is an appropriate theoretical framework to explain the formation of teachers' intentions to formative assessment. Instrumental attitude, subjective norm, and self-efficacy were found to be significant predictors of teachers' intentions to conduct formative assessment. Teachers who had a favourable instrumental attitude, a positive subjective norm, and a high level of self-efficacy were more likely to have the intention to conduct formative assessment. However, the proposed TPB-based model did not explain teachers' reported formative assessment practices very well. It is evident that external and contextual factors should be considered when examining teachers' formative assessment practices. More empirical research in this direction is warranted.

Acknowledgements

This research was partially supported by the Internal Research Grant (grant no. RG 3/2013-2014R) from the Hong Kong Institute of Education.

Appendix 1

Teachers' Conceptions and Practices of Formative Assessment Questionnaire.

The following are statements about *Formative Assessment*. Based on your understanding and teaching situations, please indicate the degree of your agreement with each statement on the answer sheet.

Formative Assessment refers to assessment activities undertaken by teachers and/or by their students in daily teaching and learning process. These assessment activities provide information to be used as feedback to modify the teaching and learning activities.

		Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
Affective Attitude Scale							
Q1	I like Formative Assessment.	1	2	3	4	5	6
Q2	Formative Assessment is an enjoyable process.	1	2	3	4	5	6
Q3	Formative Assessment is interesting.	1	2	3	4	5	6
Q4	Formative Assessment makes my teaching easier.	1	2	3	4	5	6
Q5	Formative Assessment encourages students to help each other.	1	2	3	4	5	6
Q6	Formative Assessment facilitates a better learning atmosphere.	1	2	3	4	5	6
Q7	Formative Assessment is worthy of my effort.	1	2	3	4	5	6
Instrumental Attitude Scale							

(*continued*)

		Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
Q8	Formative Assessment can raise students' interest in learning.	1	2	3	4	5	6
Q9	Formative Assessment can offer an accurate appraisal of students' performance.	1	2	3	4	5	6
Q10	Formative Assessment can integrate learning and teaching with assessment.	1	2	3	4	5	6
Q11	Formative Assessment encourages students to work harder.	1	2	3	4	5	6
Q12	Formative Assessment can offer a fair appraisal of students' performance.	1	2	3	4	5	6
Q13	Formative Assessment helps students to understand their strengths and weaknesses through feedback from teachers.	1	2	3	4	5	6
Q14	Formative Assessment can encourage autonomous learning of students.	1	2	3	4	5	6
Q15	Formative Assessment can improve student' confidence in learning.	1	2	3	4	5	6
Q16	Formative Assessment can improve the quality of teaching and learning.	1	2	3	4	5	6
Q17	Formative Assessment can improve teaching efficiency.	1	2	3	4	5	6
Subjective Norm Scale							
As far as I know, the following stakeholders believe that Formative Assessment should be implemented.							
Q18	Officials of the Education Bureau	1	2	3	4	5	6
Q19	The principal of my school	1	2	3	4	5	6
Q20	Parents of my students	1	2	3	4	5	6
Q21	My students	1	2	3	4	5	6
Q22	My colleagues	1	2	3	4	5	6
Controllability Scale							
Q23	I can decide the frequency of implementing Formative Assessment.	1	2	3	4	5	6
Q24	I can decide the timing of implementing Formative Assessment.	1	2	3	4	5	6
Q25	I can decide whether or not to implement Formative Assessment.	1	2	3	4	5	6
Q26	I can decide the method of the implementation of Formative Assessment.	1	2	3	4	5	6
Self-Efficacy Scale							
Q27	I can integrate Formative Assessment into the teaching and learning process.	1	2	3	4	5	6
Q28	I have received sufficient training to implement Formative Assessment.	1	2	3	4	5	6
Q29	I can design appropriate assessment tasks for Formative Assessment.	1	2	3	4	5	6
Q30	I have enough time to implement Formative Assessment.	1	2	3	4	5	6
Q31	I have sufficient supporting materials (e.g., handbook, DVD) to implement Formative Assessment.	1	2	3	4	5	6
Q32	I have sufficient skills to implement Formative Assessment.	1	2	3	4	5	6
Intention Scale							
Q33	I am willing to try to implement Formative Assessment.	1	2	3	4	5	6
Q34	I am willing to integrate Formative Assessment into my teaching.	1	2	3	4	5	6
Q35	I am willing to design appropriate assessment tasks for Formative Assessment.	1	2	3	4	5	6
Q36	I am willing to adjust the assessment methods to meet the requirements of Formative Assessment.	1	2	3	4	5	6
Q37	I am willing to make effort to implement Formative Assessment.	1	2	3	4	5	6
Q38	I am willing to encourage students to participate in Formative Assessment.	1	2	3	4	5	6
Behaviour Scale							
Q39	In the past six months, how often have you implemented Formative Assessment?						
	☐1 Everyday			☐2 Almost everyday		☐3 Most days	
	☐4 A number of days, but less than half			☐5 Some days		☐6 Never	
Q40	In the past six months, please estimate how frequent you have implemented Formative Assessment in your teaching?						
	☐1 Very frequent			☐2 Frequent		☐3 Sometimes	
	☐☐4 Seldom			☐5 Rarely		☐6 Never	

References

Ajzen, I. (1985). From intentions to actions: a theory of planned behavior. In J. Kuhl, & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). Berlin: Springer Berlin Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Ajzen, I. (2002a). *Constructing a TpB questionnaire: Conceptual and methodological considerations*. Retrieved from http://www.unibielefeld.de/ikg/zick/ajzen%20construction%20a%20tpb%20questionnaire.pdf.

Ajzen, I. (2002b). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32*(4), 665–683.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557.

Arbuckle, J. L. (2011). *IBM SPSS Amos 20 user's guide*. Armonk, NY: IBM.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: a meta-analytic review. *British Journal of Social Psychology, 40*(4), 471–499.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman and Company.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*(5), 536–553.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp. 9–25). London: Sage Publications.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301–318.

Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: a tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5/6), 307–320.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347–363.

Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction, 7*(1), 107–120.

Courneya, K. S., Vallance, J. K., Jones, L. W., & Reiman, T. (2005). Correlates of exercise intentions in non-Hodgkin's lymphoma survivors: an application of the theory of planned behavior. *Journal of Sport and Exercise Psychology, 27*(3), 335–349.

Curriculum Development Council (CDC). (2001). *Learning to learn – The way forward in curriculum development*. Hong Kong: Educationt Department.

Curriculum Development Council and the Hong Kong Examinations and Assessment Authority (CDC & HKEAA). (2007). *Physical education curriculum and assessment guide (Secondary 4–6)*. Hong Kong: Education Department.

Davis, L. E., Ajzen, I., Saunders, J., & Williams, T. (2002). The decision of African American students to complete high school: an application of the theory of planned behaviour. *Journal of Educational Psychology, 94*(4), 810–819.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1–38.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development, 13*(2), 173–186.

Dodgson, J. E., Henly, S. J., Duckett, L., & Tarrant, M. (2003). Theory of planned behavior-based models for breastfeeding duration among Hong Kong mothers. *Nursing Research, 52*(3), 148–158.

Duwairi, A. A. (2013). Secondary teachers' conceptions and practices of assessment models: the case for mathematics teachers in Jordan. *Education, 134*(1), 124–133.

Garson, G. D. (2009). *Structural equation modeling*. Retrieved from http://www.statisticalassociates.com/sem.htm.

Hallinger, P. (2011). Making education reform happen: Is there an "Asian" way? *School Leadership and Management, 30*(5), 401–418.

Harrison, C. (2013). Collaborative action research as a tool for generating formative feedback on teachers' classroom assessment practice: the KREST project. *Teachers and Teaching: Theory and Practice, 19*(2), 202–213.

Harrison, C., & Howard, S. (2009). *Inside the primary black box: Assessment for learning in primary and early years classrooms*. London: GL Assessment.

Hart, D. L., & Wright, B. D. (2002). Development of an index of physical functional health status in rehabilitation. *Archives of Physical Medicine and Rehabilitation, 83*(5), 655–665.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.

Hay, P. J. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 312–325). London: Sage Publications.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Kersaint, G., Lewis, J., Potter, R., & Meisels, G. (2007). Why teachers leave: factors that influence retention and resignation. *Teaching and Teacher Education, 23*(6), 775–794.

Lawton, R., Conner, M., & Parker, D. (2007). Beyond cognition: predicting health risk behaviours from instrumental and affective beliefs. *Health Psychology, 26*(3), 259–267.

Linacre, J. M. (2006). *A user's guide to WINSTEPS/MINISTEP: Rasch-model computer programs*. Chicago, IL: Winsteps.

Linacre, J. M. (2011). *Winsteps® (Version 3.7) [Computer Software]*. Beaverton, Oregon: Winsteps.com. Available from http://www.winsteps.com/.

Lipnevich, A. A., MacCann, C., Krumm, S., Burrus, J., & Roberts, R. D. (2011). Mathematics attitudes and mathematics outcomes of US and Belarusian middle school students. *Journal of Educational Psychology, 103*(1), 105–118.

MacFarlane, K., & Woolfson, L. M. (2013). Teacher attitudes and behaviour toward the inclusion of children with social, emotional and behavioural difficulties in mainstream schools: an application of the theory of planned behaviour. *Teaching and Teacher Education, 29*, 46–52.

Maclellan, E. (2004). Initial knowledge states about assessment: novice teachers' conceptualisations. *Teaching and Teacher Education, 20*(5), 523–535.

Mok, W. K., & Lee, A. Y. K. (2013). A case study on application of the theory of planned behaviour: predicting physical activity of adolescents in Hong Kong. *Journal of Community Medicine & Health Education, 3*(5). http://dx.doi.org/10.4172/2161-0711.1000231.

Mulholland, J., & Wallace, J. (2001). Teacher induction and elementary science teaching: enhancing self-efficacy. *Teaching and Teacher Education, 17*(2), 243–261.

Ofsted. (2007). *Annual report of her majesty's chief inspector, children's services and skills 2006–2007*. London: Her Majesty's Stationery Office.

Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: rasch measurement. *British Educational Research Journal, 36*(4), 611–626.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement test*. Copenhagan, Denmark: Danish Institute for Educational Research. Expanded ed. (1980). Chicago, IL: The University of Chicago Press.

Rhodes, R. E., & Courneya, K. S. (2003). Investigating multiple components of attitude, subjective norm, and perceived control: an examination of the theory of planned behaviour in the exercise domain. *British Journal of Social Psychology, 42*(1), 129–146.

Rink, J., & Mitchell, M. (2002). High stakes assessment: a journey into unknown territory. *Quest, 54*(3), 205–223.

Rink, J., & Williams, L. (2003). Developing and implementing a state assessment program. *Journal of Teaching in Physical Education., 22*(5), 473–493.

Smith, K. (2013). Formative assessment of teacher learning: Issues about quality, design characteristics and impact on teacher learning. *Teachers and Teaching: theory and practice, 19*(2), 228–234.

Tan, A. L., & Towndrow, P. A. (2009). Catalyzing student – teacher interactions and teacher learning in science practical formative assessment with digital video technology. *Teaching and Teacher Education, 25*(1), 61–67.

Törmäkangas, K. (2011). Advantages of the Rasch measurement model in analysing educational tests: an applicator's reflection. *Educational Research and Evaluation: An International Journal on Theory and Practice, 17*(5), 307–320.

Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioral control is a multidimensional construct: perceived control and perceived difficulty. *British Journal of Social Psychology, 41*, 101–121.

Trafimow, D., Sheeran, P., Lombardo, B., Finlay, K. A., Brown, J., & Armitage, C. J. (2004). Affective and cognitive control of persons and behaviours. *British Journal of Social Psychology, 43*(2), 207–224.

Yan, Z. (2014). Predicting teachers' intentions to implement school-based assessment using the theory of planned behaviour. *Educational Research and Evaluation, 20*(2), 83–97.

Yan, Z., & Sin, K. F. (2014). Inclusive education: teachers' intentions and behaviour analysed from the viewpoint of the theory of planned behaviour. *International Journal of Inclusive Education, 18*(1), 72–85.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65.

Wong, M. W. Y. (2014). Assessment for learning, a decade on: self-reported assessment practices of secondary school music teachers in Hong Kong. *International Journal of Music Education, 32*(1), 70–83.

Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice, 16*(4), 33–45.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.