Teaching and Teacher Education 114 (2022) 103718

Contents lists available at ScienceDirect

Teaching and Teacher Education

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

Research paper

Predicting teachers' formative assessment practices: Teacher personal and contextual factors

Zi Yan [a, *], Ming Ming Chiu [b], Eric Chi Keung Cheng [a]

[a] Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong
[b] Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong

HIGHLIGHTS

- Person and context factors affect teachers' formative assessment practices (FAP).
- Teachers with more self-efficacy reported more FAP.
- Teachers in schools with more teachers or more school support reported more FAP.
- Teachers with higher rank or less teaching experience reported more FAP.

ARTICLE INFO

Article history:
Received 9 November 2021
Received in revised form
13 February 2022
Accepted 21 March 2022
Available online 9 April 2022

Keywords:
Formative assessment
Teacher
Personal factor
Contextual factor

ABSTRACT

Past studies identified isolated factors influencing teachers' formative assessment practices (FAP) but did not show a comprehensive model of it. Hence, this study systematically creates and tests a theoretical model of FAP by determining whether personal or contextual factors are linked to teachers' FAP. We collected longitudinal data from 296 Hong Kong primary and secondary teachers through two waves of the survey. The results of a multivariate outcome, multilevel analysis showed that personal (teacher rank, years of teaching experience, and self-efficacy) and contextual factors (school support, total number of teachers in a school, and teaching grade) significantly impact teachers' FAP.

1. Introduction

By informing instruction and scaffolding student learning, formative assessment can help drive education reforms (e.g., Assessment Reform Group, UK; No Child Left Behind, US). Empirical studies have shown that formative assessment improves students' learning motivation (Weurlander et al., 2012), self-regulated learning (Clark, 2012; Ng, 2016), and learning outcomes (Black & Wiliam, 1998; Klute et al., 2017; Wiliam et al., 2004). How teachers perceive, design, and implement formative assessment activities in their classrooms strongly affects its success, but many teachers are reluctant to change their conceptions and practices to implement formative assessment (Boardman & Woodruff, 2004; Brookhart, 2011). As many teachers cannot easily implement it (Ali & Iqbal, 2013; Cowie & Harrison, 2016), they often avoid doing it or do it poorly (DeLuca & Bellara, 2013; Wylie & Lyon, 2015). Since Hong Kong policies have advocated formative assessment for two decades, her teachers are well aware of its pedagogical merits, but its implementation in classrooms is not satisfactory (Yan & Brown, 2021). Past studies identified isolated factors influencing teachers' formative assessment practices (FAP) but did not test a systemic model of them. Hence, instead of examining only one factor or a few factors, this study takes steps toward a comprehensive theory of how personal and contextual factors simultaneously influence teachers' FAP. The findings have the potential to inform the deployment of suitable measures to support FAP implementation.

1.1. Teachers' FAP

Although scholars generally agree that formative assessment involves eliciting evidence about student learning and using that evidence to make decisions about future teaching or learning plans

* Corresponding author. Department of Curriculum and Instruction, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, N.T., Hong Kong.
E-mail address: zyan@eduhk.hk (Z. Yan).

https://doi.org/10.1016/j.tate.2022.103718

(Black & Wiliam, 2009; Clark, 2012), formative assessment remains an umbrella term for various practices (Bennett, 2011; McMillan et al., 2013). To provide a unifying basis for FAP, Wiliam and Thompson (2008) argued that all FAPs include one or more of these five strategies:

1) Clarifying and sharing learning intentions and criteria for success;
2) Engineering effective classroom discussions, questions, and learning tasks;
3) Providing feedback that aids learners' progress;
4) Activating students as instructional resources for one another; and
5) Activating students as the owners of their learning.

This unifying basis covers a wide range of essential formative assessment aspects, including recognition of students' active role in formative assessment (Allal, 2016).

1.2. Predictors of teachers' FAP

Personal or contextual factors can affect teachers' FAP. Such personal factors include knowledge or beliefs (Brown et al., 2011), while such contextual factors include accountability, school culture, and school leadership/management (Brookhart & Moss, 2013; McKay, 2006; Sach, 2015). By understanding the facilitators and impediments of teachers' FAP, educators can develop suitable supports to empower facilitators and overcome impediments to enact and maximise FAP.

Three major review studies have synthesised the main findings of FAP's antecedents (see summary in Table 1, organised into *personal* and *contextual* factors, along with factors examined in this study). Using Kozma's (2003) three-level (teacher, school/community, and society/culture) framework, Fulmer et al. (2015) show that the following factors influenced teachers' assessment practices (including formative assessment): (a) teacher's student grade level (kindergarten, primary, etc.), subject matter, teaching experience, managerial responsibility, values, conceptions, and knowledge regarding assessment practice, (b) school leaders, community, and school climate, and (c) national curriculum, cultural norms, and national policies. Later, Heitink et al. (2016) found that the following aspects affect the successful implementation of formative assessment (he used the term "assessment for learning"): assessment design; both teachers' and students' knowledge, skills, beliefs, and attitudes regarding formative assessment; school leaders, school culture and support, and professional development. More recently, Yan et al. (2021) found that personal and contextual factors influence teachers' formative assessment intentions and practices: (a) education/training, instrumental attitude, belief of teaching, skill and ability, self-efficacy, subjective norm, and (b) school environment, internal school support, working condition, student characteristics, external policy, and cultural norms.

1.3. Personal factors

Previous studies have examined the effect of demographic factors on assessment-relevant practices, but not formative assessment. Teachers who taught for more years than other teachers used alternative assessments more often in Bol et al. (1998) and preferred original, teacher-created assessment activities over practical, easy-to-use assessment tools in Unal and Unal (2019), but years of experience and ranks/titles were unrelated to their assessment literacy in Xu and Brown (2017). As teachers at higher ranks are often more experienced, have better teaching skills and receive higher salaries than teachers at lower ranks, the former are more likely than the latter to use FAP, an integral part of teaching (Black & Wiliam, 1998; Moss & Brookhart; 2019).

H-1a. Teachers with more years of experience use FAP more often

vs H-1b. Teachers with more years of experience use FAP less often.

H-2a. Teachers at higher ranks use FAP more often

vs H-2b. Teachers at higher ranks use FAP less often.

Teachers' beliefs, attitudes, knowledge, and skills can affect their FAP (Black & Wiliam, 1998; Pastore & Andrade, 2019). Teachers' beliefs about teaching and learning affect their understanding of the purposes and use of assessment (Harrison, 2013; Karim, 2015; Yan, 2014). According to the Theory of Planned Behaviour (TPB, Ajzen, 1991), personal attitude, self-efficacy, and subjective norm affect behaviour intention formulation. Behaviour intention and self-efficacy predict actual behaviour. Teachers with positive instrumental attitudes towards formative assessment view it as a useful tool to enhance teaching effectiveness and improve students' learning outcomes, so they are more likely to implement FAP (Karaman & Sahin, 2017; Yan & Cheng, 2015). Teachers with greater confidence in their assessment competence (self-efficacy) are also more likely than other teachers to both intend to and actually implement FAP (Crichton & McDaid, 2016; So & Lee, 2011). Likewise, perceived positive opinions from important stakeholders (subjective norm) are a social impetus for teachers to use FAP more often (Yan & Cheng, 2015). Hence, we test the following hypotheses.

H-3a. Among teachers, those with better instrumental attitudes use FAP more often.

H-3b. Among teachers, those with subjective norms supporting FAP use FAP more often.

H-3c. Among teachers, those with greater intention to use FAP more often.

Table 1
F*actors influencing teachers' assessment practices.*

Studies	Target practice	Personal Factors	Contextual Factors
Fulmer et al. (2015)	Assessment in general	Teaching experience, managerial responsibility, values, conceptions, and knowledge	School sector, subject matter, school leaders, community, school culture and climate, national curriculum, cultural norms, and national or international policies
Heitink et al. (2016)	Assessment for learning	Knowledge, skills, beliefs, and attitudes	School leaders, school culture, school support, professional development, student knowledge and skill, student beliefs and attitude
Yan et al. (2021)	Formative assessment	Education and training, instrumental attitude, belief of teaching, skill and ability, self-efficacy, affective attitude, and subjective norm	School environment, school support, working conditions, student characteristics, external policy, and cultural norms
The current study	Formative assessment	Years of teaching, rank, self-efficacy, instrumental attitude, subjective norm, intention	School context: school banding, total number of teachers, examination culture, school support, teaching environment, student attributes Course context: teaching grade, subject

H-3d. Among teachers, those with greater self-efficacy use FAP more often.

1.4. Contextual factors

As FAP occurs in a complex environment, course/school/social factors can affect FAP. Fulmer et al. (2015) argued that teachers' perceptions and implementation of assessment practices might be affected by different levels of contextual factors. Those factors may come from the immediate course context (e.g., subject area taught and teaching grade), the school and surrounding community (e.g., student characteristics, school support, teaching environment), or macro-social factors (e.g., exanimation culture). For example, teaching subject might influence teachers' assessment practices. Zhang and Burry-Stock (2003) found that teachers of academic subjects (e.g., mathematics, language, and science) engaged in certain assessment activities (e.g., interpreting and communicating standardised test results) more often than teachers of non-academic subjects (e.g., arts, music, and physical education). Bol et al. (1998) showed that mathematics teachers used traditional assessment methods (e.g., examination) less frequently than teachers of science, social studies, and English. They speculated that the emphasis on problem-solving in mathematics encourages mathematics teachers to understand students' learning process. So, we propose the following hypothesis.

H-4. Mathematics teachers use FAP more often than other teachers do.

Past studies have not examined how teaching grade affects teachers' FAP, although some studies showed that compared to high school teachers, elementary school teachers used alternative assessment strategies more often (Bol et al., 1998; Zhang & Burry-Stock, 2003). We propose that when students' performance on a high-stakes evaluation (e.g., university entrance exam) affects their teachers (e.g., salary, reputation, etc.), such teachers are more likely to use teaching methods, such as FAP, to aid their students' learning. Secondary school seniors' (S6) performance on university entrance exams affects their teachers' and schools' reputations in many education systems (including Hong Kong), so their teachers might use FAP more often than other teachers do.

H-5. S6 teachers use FAP more often than other teachers do.

The total number of teachers and school banding might be linked to FAP. Schools with more teachers are more likely to include some teachers with non-traditional teaching methods and assessment strategies, such as FAP. More such teachers both increases the likelihood that some teachers will use it and that such teachers mutually support one another's use within a FAP-supportive culture (*social contagion*, Christakis & Fowler, 2013).

H-6. Among teachers, those in schools with more teachers use FAP more often.

Students with higher achievement (*school banding*) or better attitudes toward formative assessment, better understanding and readiness for formative assessment (*student attributes*) are more likely to adapt readily to new teaching or assessment methods, which facilitates teacher use of FAP (Grob et al., 2017; Heitink et al., 2016; Yan et al., 2021). Hence, teachers in such schools have better perceptions of formative assessment (van der Kleij, 2019) and might be more willing to use FAP.

H-7a. Among teachers, those in schools with higher-achieving students (indicated by higher school banding) use FAP more often.

H-7b. Among teachers, those perceiving better student attributes among their students use FAP more often.

Teachers' perceptions of school support and teaching environment can also affect FAP (Fulmer et al., 2015; Heitink et al., 2016). School support refers to school leadership, policies and resources that facilitate FAP. School leaders who give more FAP encouragement and support (a) signal the importance of FAP, (b) help teachers successfully apply FAP, and (c) lay the groundwork for continued sustained use of FAP (Moss et al., 2013; Sach, 2015). Hence, such schools often have teachers with more positive attitudes toward intended and actual FAP (Brink & Bartz, 2017; Brookhart & Moss, 2013).

H-8a. Among teachers, those perceiving greater school support use FAP more often.

Teaching environment is the immediate environment, including important conditions (e.g., instructional time, curricular requirement, workload, etc.) that influence FAP. When teachers perceive sufficient instructional time, little pressure to cover a broad curriculum, and a less workload, they are more likely to use FAP (Crichton & McDaid, 2016; Moss et al., 2013). By contrast, teachers with insufficient instruction time, substantial pressure to cover a broad curriculum, or a heavy workload, the excessive demands reduce their likelihood of using FAP.

H-8b. Among teachers, those perceiving better teaching environments use FAP more often.

Examination culture is a social norm that sees assessment as a tool of accountability and a standard of achievement. In such a high-stakes assessment culture, teachers value summative assessments and prioritise students' high learning outcomes, as measured via grades (Deneen et al., 2019). As FAP requires substantial time, it can compete with coverage of more subject matter content on the high-stakes exams in both the Eastern (Yan & Brown, 2021) and Western education systems (Bonner, 2016; Wiliam et al., 2004). Hence, examination culture might be negatively linked to FAP.

H-9. Among teachers, those who perceive a greater examination culture use FAP less often.

1.5. This study

The following overarching research question guided this study: what personal or contextual factors are linked to the frequency of teachers' FAP? This study advances this line of research in two ways. First, as Yan et al.'s (2021) review suggests, a comprehensive model of teachers' FAP should integrate personal and contextual factors rather than treating them as two isolated segments. Hence, we seek empirical evidence about how a systemic model of factors is linked to teachers' FAP and explore relations among these factors, including indirect mediation effects and interaction (moderation) effects. Second, this study collects longitudinal data to better explore causal relations. As we test whether predictors affect subsequent outcomes, we collect data on predictors before collecting them on outcomes (Sussman & Gifford, 2019; Yan & Sin, 2014). The longitudinal design improves the testing accuracy of the relations between predictors and target behaviour in the proposed model.

1.6. The setting

Since Hong Kong Education Commission's (2000) proposal, "Learning for Life – Learning through Life", an assessment reform has been initiated in Hong Kong. This reform promoted formative assessments in the teaching and learning circle to contrast with the focus on summative examinations. The Hong Kong government has invested significant resources in the reform, such as integrating formative assessment into school-based policies, providing professional development, facilitating diversified modes of assessment in classrooms, and so on. Nevertheless, the assessment culture in Hong Kong has not been really changed. High-stakes examinations are still used to make important decisions about students' futures (e.g., students are selected for government places in universities based on their grades in the Hong Kong Diploma of Secondary

Education examination). Although Hong Kong teachers hold a generally positive attitude towards formative assessment (Brown et al., 2009; Hui, 2012), the frequency and quality of formative assessment in classrooms are still unsatisfactory (Lam, 2018; Yan & Cheng, 2015). Readers can refer to Yan and Brown (2021) for a review on the assessment reform in Hong Kong.

2. Methodology

2.1. Sample

In this study, 296 teachers from 12 Hong Kong schools (seven primary and five secondary schools) participated. The participating schools were purposefully selected to represent different levels of student academic performance (high, medium, and low). Of these, 169 (57.1%) taught primary school and 127 (42.9%) taught secondary school. The majority of the teachers are female (n = 192, 64.9%). Mean teaching experience was 13.4 years (SD = 9.7), ranging from 1 to 37 years.

Statistical power differs across levels. For $\alpha = 0.05$ and a small effect size of 0.2, statistical power for 296 teachers is 0.93 and for 12 schools is less than 0.25 (Konstantopoulos, 2008). This sample size also exceeds the minimum requirement of 180 for a structural equation model that accounts for 45% of the variance (Wolf et al., 2013). Due to the few schools' low statistical power, the likelihood that a non-significant result for a school variable is a false negative is high, but we retain our usual confidence in our significant results (Kennedy, 2008).

2.2. Instrument

Appropriate instruments for assessing the frequency of teachers' FAP are surprisingly scarce. Some studies on teachers' formative assessment practices created and used instruments without formally validating them (e.g., McMillan et al., 2010; Song & Koh, 2010). Some researchers developed instruments relevant to formative assessment, but may not be appropriate for this study. For example, DeLuca et al. (2018) developed the Approaches to Classroom Assessment Instrument (ACAI) which provides an overall picture of teachers' "approach to classroom assessment", but has no subscale focusing on teachers' formative assessment practices. Furthermore, ACAI's length (70 items) limits its use of large samples.

Thus, this study used the Teacher Formative Assessment Practice Scale (TFAPS) (Yan & Pastore, 2022) to assess the frequency of teachers' FAP. The TFAPS is a theory-driven, parsimonious scale (10 items; Cronbach's $\alpha = 0.77$; e.g., I point out students' strengths and weaknesses in my feedback.) that covers all five major strategies in Wiliam and Thompson's (2008) framework of formative assessment (as discussed earlier). This Likert-type scale ranges from Never (1) to Very frequently (6).

Four predictors of formative assessment were assessed with the Teacher's Conceptions and Practices of Formative Assessment Questionnaire (Yan & Cheng, 2015): (a) instrumental attitude (10 items; Rasch reliability = 0.88; e.g., Formative assessment can encourage autonomous learning of students.), self-efficacy (6 items; Rasch reliability = 0.84; e.g., I have sufficient skills to implement formative assessment.), subjective norm (5 items; Rasch reliability = 0.75; e.g., The principal of my school believes that formative assessment should be implemented.), and intention (6 items; Rasch reliability = 0.88; e.g., I am willing to integrate formative assessment into my teaching.). This Likert-type scale ranges from Strongly Disagree (1) to Strongly Agree (6).

In addition, four new scales were developed in this study to assess four contextual factors, respectively, including *examination culture, teaching environment, school* support, *and student attribute.* The items for each contextual factor were developed based on the research literature and three focus group interviews with teachers ($N = 12$). The scale development followed Onwuegbuzie et al.'s (2010) Instrument Development and Construct Validation (IDCV) process informed by DeVellis' (2012) recommendations for scale development. Starting with a construct map that defined the constructs, the research team then described behaviours/indicators that undergird the constructs and accordingly built an item pool based on the literature and the results of the focus group interviews. Next, experts with formative assessment expertise and familiarity with the teaching environment in Hong Kong schools reviewed the items for face validity, content validity, readability, relevance, and possible bias. The draft questionnaire was piloted on a small convenience sample of 50 teachers. Next, we categorised and scored their responses to each item and incorporated their written feedback to further improve the items. Quantitative and qualitative analyses of the items informed further modifications. Each of the developed scales has five items: examination culture (e.g., To me, helping students get high scores on examinations is more important than doing formative assessment), teaching environment (e.g., My workload is appropriate for doing formative assessment.), school support (e.g., My school provides professional support for my implementation of formative assessment.), and student attribute (e.g., My students' attitudes toward formative assessment support my implementation of formative assessment.). This Likert-type scale ranges from Strongly Disagree (1) to Strongly Agree (6).

The questionnaire for the participants begins with a definition of formative assessment to build a common understanding of it among them: assessment activities undertaken by teachers and/or by their students in daily teaching and learning processes; these assessment activities provide information to be used as feedback by teachers and/or their students to modify the teaching and learning strategies. As formative assessment activities occur in context, teachers were required to respond to all items in the context of teaching one major subject at one school grade level during the last year.

2.3. Procedure

We collected two waves of surveys. The survey at Time 1 concurrently collected personal measures, contextual measures, and teachers' FAP, so we named this FAP "*Immediate FAP*". One year later, the Time 2 survey collected teachers' FAP, which we named "*Post-FAP*". Teachers' responses at Time 1 and Time 2 were matched through an eight-digit identifier number (last four digits of his/her telephone number with the four-digit birth date [month and day of month]. We obtained ethics approval from our University. We informed participants that their participation was voluntary and that they could withdraw from the study at any time without any negative consequences. The data collected is used only for research purposes, and no identifiable information is disclosed.

2.4. Data analysis

2.4.1. Analytic issues and statistics strategies

To address analytic issues regarding data, outcomes, and explanatory variables, we used suitable statistics strategies (see Table 2). First, the data had missing data and survey measurement error. Missing data (1% of these data in this study) biased results, lowered estimation efficiency and hindered data analyses. Thus, we estimated the missing data with *Markov Chain Monte Carlo multiple imputation* (MCMC-MI). In computer simulations, MCMC-MI outperformed *pairwise deletion*, *listwise deletion, mean substitution*, and

Table 2
Statistics strategies to address each analytic difficulty.

Analytic difficulty	Statistics strategy
Data Set	
• Missing data (01??10011)	• Markov Chain Monte Carlo multiple imputation (Peugh & Enders, 2004)
• Measurement errors on surveys	• Factor Analysis (Joreskog & Sorbom, 2018) • Item Response model (Embretson & Reise, 2013)
Outcome variables	
• Nested data (Teachers within schools …)	• Multilevel analysis (aka Hierarchical linear modelling, Hox et al. (2017)
• Multiple outcomes dependent variables (Y_1, $\mathbf{Y_2}$, ...)	• Multivariate outcome multilevel analysis (Hox et al., 2017) • Multilevel Structural equation model (Joreskog & Sorbom, 2018)
Explanatory variables	
• Indirect, multilevel mediation effects (X →**M**→Y)	• Multilevel *M-test* (MacKinnon at al., 2004) • Multilevel structural equation model (Joreskog & Sorbom, 2018)
• Many hypotheses' false positives	• Two-stage linear step-up procedure (Benjamini et al., 2006)
• Cross-level interactions (Teacher x School)	• Random effects model (Hox et al., 2017)
• Interaction in structural equation model	• Residual centering (Little et al., 2012)
• Compare effect sizes ($\beta_1 > \beta_2$?)	• Lagrange multiplier tests (Bertsekas, 2014)
• Consistency of results across data sets (Robustness)	• Separate multilevel, single outcome models • Analyses of subsets of the data (Kennedy, 2008) • Original (not estimated) data

simple imputation (Peugh & Enders, 2004; specifically with *LISREL 10.1* [Joreskog & Sorbom, 2018] and these settings: [a] single-chain, [b] EM Posterior mode initial estimates, [c] Jeffreys priors, [d] 500 imputations, [e] 200 burn-in iterations, and [f] 100 iterations). To reduce survey measurement error, we used multiple questions for each construct to create a precise index. We analysed whether sets of questions reflected one or more underlying constructs (e.g., self-efficacy) via *confirmatory factor analyses* (Joreskog & Sorbom, 2018).

Outcome issues included school differences and multiple outcomes. As teachers in the same school likely resembled one another more than those in different schools (*nested data*), an ordinary least squares regression biased the *standard errors*, so we used a *multilevel analysis* (Hox et al., 2017). As we modelled multiple outcomes, correlated residuals underestimated standard errors, which we addressed via a *multivariate outcome multilevel analysis* (Hox et al., 2017) and a *multilevel structural equation model* (Joreskog & Sorbom, 2018).

Explanatory variable issues included indirect effects, many hypotheses' false positives, cross-level interactions, effect size comparisons, and robustness. Separate, single-level tests of indirect mediation effects on nested data biased results, so we tested for simultaneous multilevel mediation effects with a *multilevel M-test* (MacKinnon et al., 2004) and a *multilevel structural equation model* (Little et al., 2012). Conducting many hypothesis tests raised the likelihood of a false positive, so we reduced its likelihood with the *two-stage linear step-up procedure.* In computer simulations, it outperformed 13 other competing methods (Benjamini et al., 2006).

With nested data, incorrectly modelling interaction effects across levels (e.g., Student X Turn of Talk) biased the results, so we used a *random effects* model (Hox et al., 2017). If the regression coefficient of an explanatory variable (e.g., $\beta_{yvj} = \beta_{yv0} + f_{yvj}$) differed significantly across levels ($f_{yvj} \neq 0$?), then *cross-level moderation* existed, and we modelled the regression coefficient with structural variables (e.g., total teachers in school). As interaction terms were correlated with their component variables, they caused unstable results. To address this issue, we removed these correlations before testing for interaction effects via *residual centring* (Little et al., 2012).

Although *Wald* and *likelihood ratio* tests determined whether the effect sizes of explanatory variables differ, they did not apply at boundary points. To address this issue, we used *Lagrange multiplier tests.* These tests both applied to all the data and had more statistical power for small differences from the null hypothesis (Bertsekas, 2014).

Lastly, robustness tests determined whether the results are stable when there are small differences in the data or in the analyses (Kennedy, 2008). In a multivariate outcome model, a misspecification of any equation for any outcome caused errors in other equations. To address this issue, we modelled each outcome variable separately. We also tested separate subsets of the data. Lastly, we ran the analysis on the original data (no estimated data).

2.4.2. Factor analyses

We tested the internal validity of the survey items for each construct (e.g., self-efficacy) and minimised their measurement errors with *confirmatory factor analyses*; its Bartlett factor scores yield unbiased estimates of factor score parameters (Joreskog & Sorbom, 2018). To assess the fit of the CFA, we used the comparative fit index (CFI), Tucker–Lewis index (TLI), standardised root mean square residual (SRMR), and root mean square error approximation (RMSEA). They minimised Type I and Type II errors under many conditions in Hu and Bentler's (1999) simulations. We used two fit thresholds: good (CFI & TLI >0.95; SRMR <0.08; RMSEA <0.06) and moderate (0.90 < CFI & TLI <0.95; 0.08 < SRMR <0.10; 0.06 < RMSEA <0.10).

2.4.3. Explanatory model

We modelled each teacher's immediate FAP and post-FAP with a multivariate outcome, multilevel analysis, beginning with a variance components model to test for significant differences at each level: person and context (Hox et al., 2017). See statistics equations with the vector of variables in Appendix.

Explanatory variables were entered in sequential sets to estimate the variance explained by each set and to test for mediation effects (Kennedy, 2008). Demographics affected personal processes, which in turn could differ across contexts. Thus, we first entered a vector of **Personal_demographics** (*female, years of teaching,* rank [*CM, APSM, PSM, GM, SGM, PGM*]). A *nested hypothesis test* (χ^2 log likelihood) indicated whether each set of explanatory variables is significant (Kennedy, 2008). As omitting *non-significant* variables did not cause *omitted variable bias*, we removed them to increase precision and reduce *multicollinearity* (Kennedy, 2008).

Then, we entered **Personal_processes** (*, instrumental attitude, teacher self-efficacy, subjective norm, intention*), followed by **Context** (*band_1, band_2*, secondary (vs. primary) *total teachers, examination*

culture, school support, *teaching environment, student attributes regarding formative assessment*, *teaching grade,* subject [*Chinese, English, mathematics, liberal studies, science, technology, social sciences, arts,* against a baseline of physical education]). Lastly, we tested for interactions among these variables. School level (primary vs. secondary) and teacher gender were used as control variables.

We used multilevel mediation tests to create a multilevel path analysis (Hox et al., 2017), which served as an initial candidate for the ML-SEM (Joreskog & Sorbom, 2018). We removed non-significant variables to yield the final model. The *total effect* (TE) of an explanatory variable on the outcome was the sum of its direct effects and all of its indirect effects. (The indirect effect of explanatory variable X on outcome Y via mediator M [namely X → M →Y] was the product of the standardised parameter linking X to M multiplied by the total effect of M on Y [namely (X →M) * (M →Y)].) We also analysed residuals for influential outliers. Lastly, we performed the robustness procedures discussed above.

3. Results

3.1. Factor analysis

As the school level showed no significant variance (9% for post-FAP; 15% for immediate FAP), an ML-SEM was not needed, so a single-level SEM was run. All scales have high reliability coefficients (all R_c exceeded 0.84; see Table A1 in the Appendix). All factors fit the data well (see Table 3), with high factor loadings (mean = 0.750), small standard errors, and small uniqueness (see Table A2 in the Appendix). See summary statistics in Table 4.

3.2. Explanatory model

The SEM showed a good fit (SRMR = 0.083; CFI = 0.955; TLI = 0.951; RMSEA = 0.056; χ^2 [681] = 1271; $p < .001$; IFI = 0.955). See direct and indirect effects of explanatory variables on immediate and post-FAP in Table 5 and Fig. 1.

3.2.1. Immediate FAP

Teacher personal factors, school context, and course context were linked to immediate FAP. Teachers with more years of experience showed less immediate FAP (TE = −.137), supporting hypothesis H-1b. Teachers with greater self-efficacy reported more immediate FAP (TE = 0.390), supporting H-3d and showing the largest effect size for immediate FAP. In schools with more teachers (TE = 0.270) or more support (TE = 0.214), teachers showed more immediate FAP, supporting H-6 and H-8a. The interaction between total teachers in a school and years of teaching experience showed a negative interaction effect on immediate FAP (TE = −0.108). Teachers of secondary six (high school seniors) used more immediate FAP (TE = 0.179), supporting H-5. These explanatory variables accounted for over 44% of the differences in immediate FAP (squared multiple correlation [SMC] = 0.442).

3.2.2. Post-FAP

Teacher personal factors, school context, and course context were linked to post-FAP. Teachers at the lowest rank of Certificated Master/Mistress (CM) reported less post-FAP (TE = −0.198), mitigated by a moderation effect for secondary school teachers (TE = 0.139) and partially support H-2a. Also, teachers with more years of teaching experience reported less immediate FAP (TE = −0.137) and then less post-FAP (TE = 0.137 * 0.591 = −0.081), supporting H-1b.

Teachers with greater self-efficacy reported more immediate FAP (TE = 0.390) and then more post-FAP (TE = 0.390 * 0.591 = 0.231), supporting H-3d. Also, teachers who reported more immediate FAP reported more post-FAP (TE = 0.591).

School context (total teachers, school support) were linked to post-FAP. Schools with more teachers were linked to more immediate FAP (TE = 0.270), and then more post-FAP (TE = 0.270 * 0.591 = 0.160), supporting H-6. Also, teachers who perceived more school support reported more immediate FAP (TE = 0.214) and then more post-FAP (TE = 0.214 * 0.591 = 0.126), supporting H-8a. Total teachers in a school showed a negative interaction effect with years of teaching experience on immediate FAP (TE = −0.108) and then post-FAP (TE = −0.108 * 0.591 = −0.064).

Course context, specifically the grade levels of P4 and S6 were linked to post-FAP. Teachers of P4 (TE = 0.123) students reported higher post-FAP. Also, teachers of S6 students reported higher immediate FAP (TE = 0.179) and then post-FAP (TE = 0.179 * 0.591 = 0.106), supporting H-5.

All other explanatory variables and interactions were not significant. This explanatory model accounted for nearly 44% of the variance in post-FAP (SMC = 0.437). Analysis of residuals showed no substantial outliers. Robustness tests on data subsets, single outcomes, and on the original data showed similar results.

4. Discussion

This study contributes to a comprehensive theory of how personal and contextual factors simultaneously influence teachers'

Table 3
Goodness of fit measures for congeneric confirmatory factor analysis.

Factor	#	SRMR	CFI	IFI	TLI	RMSEA	χ^2	df	*p*	AGFI	RFI
Post-FAP	10	.079	.979	.980	.970	.057	60.3	31	.001	.976	.940
Immediate FAP	10	.080	.990	.990	.981	.060	47.7	23	.002	.979	.963
Examination culture	5	.044	.984	.985	.969	.058	9.9	5	.079	.984	.939
Teaching environment	5	.029	.998	.998	.996	.032	5.2	4	.269	.991	.983
School support	5	.039	.992	.992	.979	.077	11.0	4	.027	.982	.968
Student attributes	5	.057	.990	.990	.974	.097	15.1	4	.005	.978	.966
Instrumental attitude	8	.075	.991	.991	.984	.045	25.6	16	.060	.983	.957
Self-efficacy	6	.041	.999	.999	.998	.026	10.8	9	.292	.992	.987
Subjective Norm	4	.005	1.000	1.000	1.000	.000	0.1	1	.759	1.000	.999
Intention	5	.061	1.000	1.000	1.000	.032	6.5	5	.257	.997	.998

NOTE: # = number of variables; SRMR = standardised root mean square residual; CFI = comparative fit index; IFI = incremental fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error approximation; df = degrees of freedom; AGFI = adjusted goodness of fit index; RFI = relative fit index. Each set of variables for a factor was modelled separately (congeneric).

Table 4
Summary statistics (N = 296).

Variable	Mean	SD	Min	Median	Max
Post-FAP	4.147	0.681	1.669	4.204	6.000
Immediate FAP	4.232	0.597	2.610	4.261	5.774
Examination culture	3.895	0.779	1.379	3.947	5.683
Teaching environment	3.723	0.811	1.000	3.847	5.742
School support	4.194	0.671	1.541	4.131	6.000
Students attributes	4.031	0.766	1.131	4.000	5.934
Instrumental attitude	4.269	0.602	1.654	4.273	5.787
Self-efficacy	4.192	0.709	1.133	4.044	6.000
Subjective norm	4.455	0.699	1.000	4.502	6.000
Intention	4.618	0.594	2.287	4.828	6.000
Secondary	0.429	0.496	0.000	0.000	1.000
School band 1	0.331	0.471	0.000	0.000	1.000
School band 2	0.351	0.478	0.000	0.000	1.000
Mean class size	26.801	2.693	23.000	25.000	30.000
Total teachers	56.247	10.064	23.000	57.000	66.000
Female	0.655	0.476	0.000	1.000	1.000
Teacher rank					
Certificated Master/Mistress (CM)	0.054	0.227	0.000	0.000	1.000
Assistant Primary School Master/Mistress (APSM)	0.334	0.473	0.000	0.000	1.000
Primary School Master/Mistress (PSM)	0.051	0.220	0.000	0.000	1.000
Graduate Master/Mistress (GM)	0.318	0.466	0.000	0.000	1.000
Senior Graduate Master/Mistress (SGM)	0.078	0.268	0.000	0.000	1.000
Principal Graduate Master/Mistress (PGM)	0.010	0.100	0.000	0.000	1.000
Years teaching	13.318	9.664	0.000	12.000	37.000
Subject					
Chinese	0.270	0.445	0.000	0.000	1.000
English	0.253	0.436	0.000	0.000	1.000
Math	0.209	0.408	0.000	0.000	1.000
Liberal studies	0.084	0.279	0.000	0.000	1.000
Science	0.044	0.205	0.000	0.000	1.000
Technology	0.030	0.172	0.000	0.000	1.000
Social sciences	0.081	0.273	0.000	0.000	1.000
Arts	0.057	0.233	0.000	0.000	1.000
Physical education	0.041	0.198	0.000	0.000	1.000
Grade					
Teach Primary 1	0.115	0.319	0.000	0.000	1.000
Teach Primary 2	0.182	0.387	0.000	0.000	1.000
Teach Primary 3	0.081	0.273	0.000	0.000	1.000
Teach Primary 4	0.118	0.323	0.000	0.000	1.000
Teach Primary 5	0.111	0.315	0.000	0.000	1.000
Teach Primary 6	0.142	0.350	0.000	0.000	1.000
Teach Secondary 1	0.098	0.298	0.000	0.000	1.000
Teach Secondary 2	0.061	0.239	0.000	0.000	1.000
Teach Secondary 3	0.074	0.263	0.000	0.000	1.000
Teach Secondary 4	0.111	0.315	0.000	0.000	1.000
Teach Secondary 5	0.135	0.342	0.000	0.000	1.000
Teach Secondary 6	0.162	0.369	0.000	0.000	1.000

Table 5
Direct, indirect, and total effects of explanatory variables on immediate and post-formative assessment practice.

	Formative Assessment Practice			
	Immediate	Post		
Explanatory variable	Total (Direct)	Direct	Indirect	Total
Teacher personal				
Rank CM		−0.198		−0.198
Rank CM * Secondary	0.139			
Years teaching	−0.137		−0.081	−0.081
Self-efficacy	0.390		0.231	0.231
School context				
School support	0.214		0.126	0.126
Total teachers	0.270		0.160	0.160
Total teachers * Years teaching	−0.108		−0.064	−0.064
Course context				
Teach P4		0.123		0.123
Teach S6	0.179		0.106	0.106
Assessment Practice				
Immediate Formative Assessment Practice		0.591		0.591
Total explained variance	.442			.438

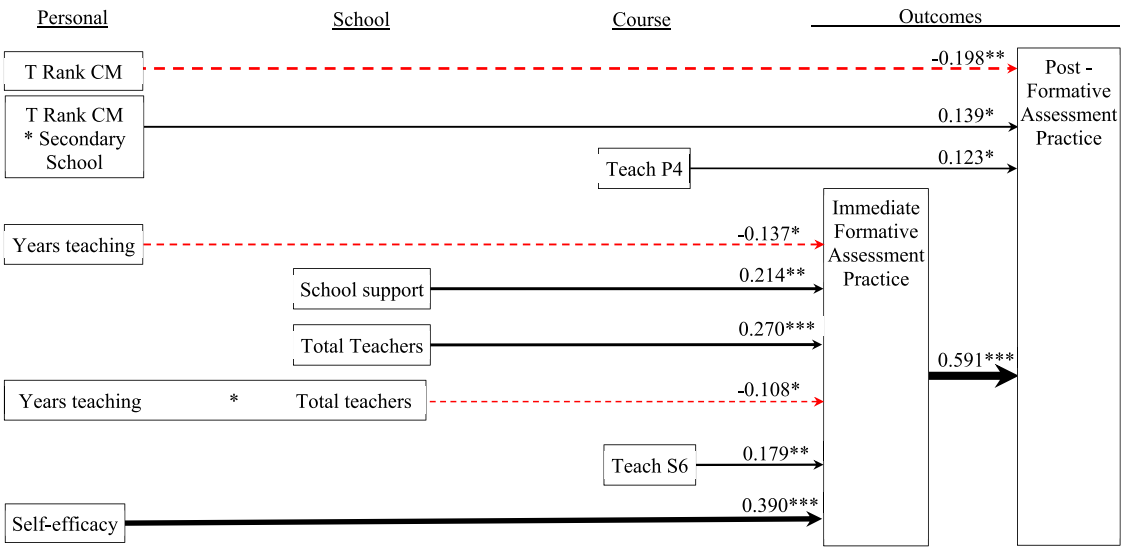

Fig. 1. Effects of explanatory variables on immediate and post-formative assessment practices.

FAP. Teachers at higher ranks, with less teaching experience, or with greater self-efficacy use FAP more often (personal factors). Also, teachers in schools with more teachers or more FAP support use FAP more often (school context). Furthermore, teachers of secondary school seniors or of primary school 4th graders showed more FAP (course context).

Despite the theoretical argument and empirical evidence supporting the use of formative assessment to enhance student learning (Andrade & Heritage, 2018; Black & Wiliam, 1998; 2009; Clark, 2012), implementing teachers' positive beliefs and knowledge about formative assessment in real classrooms is a global challenge. Past studies from different countries (Boardman & Woodruff, 2004; Brookhart, 2011; DeLuca & Bellara, 2013; Remesal, 2007; Wylie & Lyon, 2015) have shown that many teachers are reluctant to conduct formative assessment for complex reasons (Brookhart, 2011; Yan et al., 2021). This study identified the potential predictors of teachers' FAP and, therefore, shed light on how to promote teachers' FAP in classrooms.

4.1. Personal factors

Years of teaching and teacher rank were linked to FAP. Teachers with fewer years of teaching experience reported more immediate FAP and post-FAP, supporting hypothesis H-1b and consistent with the views that such teachers are more likely to be trained in the latest FAP or more open to experimenting with novel assessment methods, such as FAP. Graham et al. (2020) revealed a decline in the teaching quality after the first three years for some teachers and they attributed this decline to the removal of initial support and an increase in workload and responsibilities. Such reasons might account for less implementation of formative assessment by teachers with more years of experience.

Teachers at the lowest rank reported less post-FAP, partially supporting H-2a. However, teacher rank did not link to immediate FAP, suggesting that they might initially use FAP as much as other teachers initially but not a year later, possibly because FAP was too difficult or time demanding. Future studies can explore this speculation.

Teachers with greater self-efficacy reported more FAP, showing the strongest effect and congruent with previous findings. For example, past studies showed that self-efficacy strongly predicts teachers' intentions and/or practices regarding school-based assessment (Yan, 2014) or formative assessment (Karaman & Sahin, 2017; Yan & Cheng, 2015). Conversely, teachers with little FAP confidence were less likely to implement it (Crichton & McDaid, 2016). Even when professional training and sharing typically increased teachers' willingness to conduct FAP, they did not increase it for teachers with low FAP self-efficacy (Dixon & Haigh, 2009).

As self-efficacy substantially correlated with instrumental attitude ($r = .57$), subjective norm ($r = 0.55$), and FAP intention ($r = 0.59$), they did not separately predict FAP (H-3a, H-3b, H-3c). Furthermore, both instrumental attitude and subjective norm correlated with FAP intention respectively (r: 0.57 and 0.53). These high correlations are consistent with the Theory of Planned Behaviour (TPB, Ajzen, 1991) that attitude and subjective norm help formulate behaviour intention, which, together with self-efficacy, predicts actual behaviour. Future longitudinal studies over more time periods can rigorously test these causal link hypotheses.

4.2. School context

The many significant school and course context effects support the claim that FAP is a context-based activity (Carless, 2012; Yan et al., 2021). Teachers in schools with more teachers showed more FAP, consistent with the social contagion view that more teachers increase the likelihood that some teachers will use it and foster a mutually supportive FAP culture (Christakis & Fowler, 2013). Furthermore, Rashid and Jaidin (2014) emphasised the importance of professional sharing and discussion to facilitate FAP implementation. As total teachers in a school showed a negative interaction effect with years of teaching experience on FAP,

additional teachers with fewer years of experience are more likely to support one another's FAP.

Teachers who perceived more school support reported more immediate and post-FAP, consistent with past studies showing that encouragement and supportive measures by school leaders yielded more FAP (Brink & Bartz, 2017; Brookhart & Moss, 2013; Yan et al., 2021). As formative assessment represents a significant change from traditional assessment approaches, this result suggests that schools should institute measures to support FAP to aid its implementation. While teacher training may enhance teachers' knowledge and skills for new assessment methods, substantial change of assessment practices relies more on systematic support from a teacher's school (Yan, 2021). Without adequate school support, professional learning and collaboration among teachers can cause frustration and thereby build resistance to new assessment methods (de Jong et al., 2019), such as FAP.

However, examination culture, teaching environment, school banding, and student attributes had no significant effect on teachers' FAP (no support for H-9, H-8b, H-7a, and H-7b). Most of these teachers in Hong Kong reported a substantial examination culture (M = 3.9/6) with a small standard deviation of 0.8, so the small variation in teachers' perceived examination culture in Hong Kong might account for the lack of significant differences and the lack of a significant examination culture effect on immediate FAP or post-FAP-12.

Both teaching environment and students' formative assessment attributes correlate with school support and teachers' self-efficacy, which might account for the first two variables' non-significant effects (H-8b and H-7b). Furthermore, their correlations suggest that school support enhances a teaching environment ($r = 0.58$), which in turn enhances students' formative assessment attributes ($r_{\text{school support}} = .54$; $r_{\text{teaching environment}} = 0.50$)—all of which might support teachers' self-efficacy ($r_{\text{school support}} = .60$; $r_{\text{teaching environment}} = 0.52$; $r_{\text{students' formative assessment}} = 0.63$). Future longitudinal studies over more time periods can test these relations (H-8b, H-7b, H-3d) more rigorously.

4.3. Course context

Teachers of secondary seniors or primary four students used FAP more often than other teachers did. Teachers of secondary seniors or primary four students used FAP more often than other teachers did. Past studies have investigated the differences in assessment methods of teachers across primary and secondary schools. For example, primary school teachers used alternative assessment strategies more often than their peers in high schools (Bol et al., 1998; Zhang & Burry-Stock, 2003). However, no published study has examined the impact of teaching grade on teachers' frequency of FAP. As secondary seniors (S6) in Hong Kong take highly competitive exams for government-funded post-secondary education (enrollment slots for only 33% of students; Education Bureau, 2018), their performance affects their teachers' and high schools' reputations. So, these S6 teachers are more motivated than other teachers to help their students learn, which might account for their reports of higher immediate FAP and post-FAP. Surprisingly, teachers of P4 students also reported higher post-FAP, so future qualitative studies can explore why these teachers subsequently increased their FAP.

Teaching subject did not significantly influence teachers' FAP (no support for H-4), which differs from previous studies (e.g., Bol et al., 1998; Zhang & Burry-Stock, 2003) reporting different assessment activities across subjects. As the "assessment practices" measured in past studies and this study differ, this result should be interpreted cautiously. In past studies (e.g., Bol et al., 1998; Zhang & Burry-Stock, 2003), the "assessment practices" referred to concrete assessment activities, such as paper−pencil tests, assignments, or performance assessment. By contrast, this study measured frequency of FAP, which could include any assessment activities as long as those activities were used for formative purposes. Combining the results of past studies and this study, they suggest that teaching subject may affect teachers' choices among different assessment activities but does not influence the frequency of their FAP. Due to the small numbers of teachers, these results are exploratory and should be interpreted cautiously (e.g., 48 S6 and 35 P4 teachers in the sample; the numbers of teachers for each different academic subjects ranged from 12 [physical education] to 79 [Chinese]).

4.4. Implications for teaching and teacher professional development

In this study, the strongest predictor of teachers' FAP was their self-efficacy, followed by the total number of teachers in a school, and school support, which have the following practical implications for teaching and teacher professional development. Firstly, the strong effect of teachers' FAP efficacy in many studies suggests that it might be a critical pivot point for the uptake of FAP in classrooms. As mastery experiences can enhance self-efficacy (Bandura, 1997), teacher training that not only teaches formative assessment concepts but also equips teachers with practical strategies that aid successful FAP might substantially increase subsequent FAP in classrooms. As a proxy for mastery experiences, successful FAP examples from fellow teachers might help a new teacher visualise and enact FAP in his/her classrooms. For teachers who are struggling with formative assessment, appropriate on-site support might reduce the impact of failure experiences before establishing a sense of efficacy. Future intervention studies can test these ideas.

Secondly, formative assessment is a challenging task for teachers, so they might need their school's support for effective FAP (e.g., school leaders' public appreciation of FAP benefits, encouragement of FAP in classrooms, school policies to provide staff and professional support to teachers, and resources to facilitate classroom FAP). For example, tailor-made school-based professional development by external experts followed by internal interactions among teachers (e.g., class observation and evaluation, sharing) can enhance both teachers' knowledge and practical skills regarding formative assessment. Without strong school support, teachers might lack the inclination or resources to successfully implement FAP.

Thirdly, the effect of total teachers in a school on FAP suggests that teacher dynamics within a school affects FAP. Teachers from the same school work in the same school context, share the same resources, and face similar formative assessment difficulties, so future studies can test whether teacher training programmes making use of within-school sharing or emotional support have more influence than those exclusively relying on external parties' resources or inputs on these teachers' FAP. Compared to ad hoc

teacher training workshops, informal learning organisations, such as professional learning communities, develop teacher professional competency more sustainably (Carless, 2012). Through critical reflection and exchange of ideas in the communities, teachers can build their professional knowledge and create sustainable changes (Skerrett, 2010). Such informal learning organisations have effectively improved teachers' FAP in previous studies (e.g., Hargreaves et al., 2013).

4.5. Limitations and future research

This study's limitations included its sample, explanatory variables, and the measurement of FAP. As this study only included a convenience sample of 296 primary and secondary teachers in Hong Kong, future studies can include larger representative samples of teachers of younger or older students, and teachers in other societies. Also, this study had limited explanatory variables, so future studies can include additional explanatory variables (e.g., past professional development, current school's professional development opportunities). Furthermore, the data on teachers' FAP and its predictors were exclusively collected through self-report instruments. Thus, the results might be vulnerable to response bias (e.g. inaccurate memory and social desirability). To have a more in-depth and realistic understanding of teachers' FAP, future studies can consider using more objective measures (e.g. direct observations) to replace or substantiate self-reports. Lastly, we did not collect the teacher data on predictors at T2. As these attributes might change across time, future studies can collect and analyse these data.

5. Conclusion

This study explored whether personal or contextual factors were linked to teachers' FAP. According to the results, personal (teacher rank, years of teaching experience, self-efficacy) and contextual factors (school support, total number of teachers in a school, and teaching grade) were linked to teachers' FAP. This study takes steps toward a comprehensive theory of how personal and contextual factors simultaneously influence teachers' FAP. These findings can inform school leaders and teacher educators to develop suitable supports to empower facilitators and overcome impediments to enact and maximise FAP.

CRediT statement

All authors listed have made a substantial contribution to this paper.

Declaration of competing interest

There is no conflict of interest associated with this study.

Acknowledgements

The work described in this paper was supported by a General Research Fund from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. EDUHK 18607118). We appreciate the statistical assistance of Yik Ting Choi.

Appendix

Equations with Vectors

$$\mathbf{FAP}_{yij} = \beta_y + e_{yij} + f_{yj} \quad (1)$$

In the vector of $\mathbf{FAP}_{yij}$, outcome *y* (*immediate FAP* or *post-FAP*) of teacher *i* in context *j* has a grand mean intercept β_y, with unexplained components (*residuals*) at the person- and context-levels (e_{yij}, f_{yj}).

$$\mathbf{FAP}_{yij} = \beta_y + e_{yij} + f_{yj} + \beta_{yuj}\mathbf{Personal\_demographics}_{yij} + \beta_{yxj}\mathbf{Personal\_processes}_{yij} + \beta_{yv}\mathbf{Context}_{yj} + \beta_{yzj}\mathbf{Interactions}_{yij} \quad (2)$$

Table A1
The results of exploratory factor analysis

Factor	#	Eigenvalues					% variance explained by …		R_c	α
		1st	2nd	3rd	1st/2nd	2nd/3rd	1st	2nd		
Post-FAP	10	3.838	1.359	1.066	2.824	1.275	38%	14%	.901	.813
Immediate FAP	10	4.094	1.461	0.942	2.802	1.551	41%	15%	.863	.836
Examination culture	5	2.719	0.774	0.596	3.513	1.299	54%	15%	.844	.788
Teaching environment	5	3.006	0.779	0.522	3.859	1.492	60%	16%	.922	.829
School support	5	3.248	0.670	0.477	4.848	1.405	65%	13%	.948	.864
Student attributes	5	3.071	0.877	0.482	3.502	1.820	61%	18%	.958	.840
Instrumental attitude	8	4.581	0.863	0.582	5.308	1.483	57%	11%	.963	.893
Self-efficacy	6	4.270	0.632	0.349	6.756	1.811	71%	11%	.966	.917
Subjective Norm	4	2.604	0.715	0.415	3.642	1.723	65%	18%	.954	.819
Intention	5	3.966	0.419	0.271	9.465	1.546	79%	8%	.986	.935

NOTE: # = number of variables; R_c = reliability coefficient; α = Cronbach's alpha.

Table A2
Factor loadings, standard errors, and uniqueness from congeneric confirmatory factor analyses of variables for each construct

Variable	Standardised Factor Loadings	SE	Uniqueness
Post-FAP			
PQ1	0.364	0.062	0.867
PQ2	0.470	0.051	0.780
PQ3	0.442	0.056	0.805
PQ4	0.324	0.066	0.895
PQ5	0.592	0.049	0.650
PQ6	0.548	0.051	0.700
PQ7	0.878	0.027	0.229
PQ8	0.865	0.027	0.251
PQ9	0.755	0.037	0.431
PQ10	0.699	0.043	0.511
Immediate FAP			
Q1	0.539	0.055	0.709
Q2	0.603	0.046	0.637
Q3	0.550	0.063	0.697
Q4	0.336	0.069	0.887
Q5	0.560	0.050	0.687
Q6	0.597	0.053	0.643
Q7	0.839	0.039	0.296
Q8	0.804	0.039	0.353
Q9	0.756	0.041	0.429
Q10	0.763	0.038	0.418
Examination culture			
Q11	0.711	0.043	0.495
Q12	0.748	0.041	0.441
Q13	0.677	0.042	0.541
Q14	0.549	0.058	0.699
Q15	0.804	0.036	0.354
Teaching environment			
Q16	0.518	0.061	0.731
Q17	0.797	0.041	0.365
Q18	0.940	0.023	0.116
Q19	0.789	0.036	0.377
Q20	0.663	0.047	0.560
School support			
Q21	0.707	0.049	0.501
Q22	0.877	0.024	0.230
Q23	0.840	0.032	0.294
Q24	0.895	0.026	0.198
Q25	0.804	0.047	0.354
Student attributes			
Q26	0.545	0.064	0.703
Q27	0.736	0.036	0.458
Q28	0.815	0.036	0.335
Q29	0.974	0.028	0.052
Q30	0.734	0.051	0.461
Instrumental attitude			
Q31	0.815	0.028	0.337
Q32	0.804	0.028	0.354
Q33	0.870	0.026	0.243
Q34	0.817	0.035	0.332
Q35	0.758	0.036	0.425
Q36	0.812	0.032	0.341
Q37	0.848	0.026	0.281
Q38	0.808	0.032	0.347
Teacher Self-efficacy			
Q41	0.701	0.041	0.509
Q42	0.930	0.025	0.135
Q43	0.935	0.019	0.126
Q44	0.914	0.020	0.164
Q45	0.888	0.022	0.212
Q46	0.915	0.026	0.163
Subjective Norm			
Q54	0.773	0.063	0.402
Q55	0.782	0.041	0.389
Q56	0.821	0.045	0.326
Q57	0.863	0.031	0.255
Intention			

(*continued on next page*)

Table A2 (*continued*)

Variable	Standardised Factor Loadings	SE	Uniqueness
Post-FAP			
Q47	0.967	0.015	0.064
Q48	0.984	0.012	0.032
Q49	0.962	0.025	0.075
Q51	0.953	0.026	0.093
Q52	0.915	0.036	0.162

References

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

Ali, I., & Iqbal, H. M. (2013). Effect of formative assessment on students' achievement in science. *World Applied Sciences Journal, 26*(5), 677–687.

Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In D. Laveault, & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259–273). Springer.

Andrade, H. L., & Heritage, M. (2018). *Using formative assessment to enhance learning.* achievement: Routledge. *and academic self-regulation*.

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

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Bertsekas, D. P. (2014). *Constrained optimisation and Lagrange multiplier methods*. Academic.

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Boardman, A. G., & Woodruff, A. L. (2004). Teacher change and "high stakes" assessment: What happen to professional development? *Teaching and Teacher Education, 20*(6), 545–557.

Bol, L., Stephenson, P. L., O'connell, A. A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research, 91*(6), 323–330. https://doi.org/10.1080/00220679809597562

Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 21–39). Routledge.

Brink, M., & Bartz, D. E. (2017). Effective use of formative assessment by high school teachers. *Practical Assessment, Research and Evaluation, 22*. https://doi.org/10.7275/p86s-zc41. Article 8.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12.

Brookhart, S. M., & Moss, C. M. (2013). Leading by learning: When principals immerse themselves in learning about formative assessments and how students learn, they become better instructional leaders for teachers. *Phi Delta Kappan, 94*(8), 13.

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*, 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.

Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge.

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistical Medicine, 32*, 556–577.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205–249.

Cowie, B., & Harrison, C. (2016). Classroom processes that support effective assessment. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 335–350). Routledge.

Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: Learners' and teachers' views. *Curriculum Journal, 27*(2), 190–203. https://doi.org/10.1080/095851762015.1103278

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education, 64*(4), 356–372.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A largescale survey. *Assessment in Education: Principles, Policy & Practice, 25*(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514

Deneen, C. C., Fulmer, G. W., Brown, G. T. L., Tan, K. H. K., Leong, W. S., & Tay, H. Y. (2019). Value, practice and proficiency: Teachers' complex relationship with assessment for learning. *Teaching and Teacher Education, 80*, 39–47. https://doi.org/10.1016/j.tate.2018.12.022

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

Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development, 13*(2), 173–186. https://doi.org/10.1080/13664530903044002

Education Bureau. (2018). Task force on review of self-financing post-secondary education – review report. Author. https://www.cspe.edu.hk/resources/pdf/en/TF%20review%20report_EN.pdf.

Education Commission. (2000). *Learning for life – learning through life: Reform proposals for the education system in Hong Kong*. Author.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multilevel model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445

Graham, L. J., White, S. L. J., Cologon, K., & Piana, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*. https://doi.org/10.1016/j.tate.2020.103190. Article 103190.

Grob, R., Holmeier, M., & Labudde, P. (2017). Formative assessment to support students' competences in inquiry-based science education. *Interdisciplinary Journal of Problem-Based Learning, 11*(2). https://doi.org/10.7771/1541-5015.1673. Article 6.

Hargreaves, E., Berry, R., Lai, Y. C., Leung, P., Scott, D., & Stobart, G. (2013). Teachers' experiences of autonomy in continuing professional development: Teacher learning communities in London and Hong Kong. *Teacher Development, 17*(1), 19–34.

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. https://doi.org/10.1080/13540602.2013.741839

Heitink, M., van der Kleij, F., Veldkamp, B., Schildkamp, K., & Kippers, W. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50–62. https://doi.org/10.1016/j.edurev.2015.12.002

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

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

de Jong, L., Meirink, J., & Admiraal, W. (2019). School-based teacher collaboration: Different learning opportunities across various contexts. *Teaching and Teacher Education, 86*. https://doi.org/10.1016/j.tate.2019.102925. Article 102925.

Joreskog, K., & Sorbom, D. (2018). *Lisrel 10.1*. Scientific Software International.

Karaman, P., & Sahin, C. (2017). Adaptation of teachers' conceptions and practices of formative assessment scale into Turkish culture and a structural equation modeling. *International Electronic Journal of Environmental Education, 10*(2), 185–194. https://doi.org/10.26822/iejee.2017236114

Karim, B. H. H. (2015). The impact of teachers' beliefs and perceptions about formative assessment in the university ESL class. *International Journal of Humanities Social Sciences and Education, 2*(3), 108–115.

Kennedy, P. (2008). *Guide to econometrics*. Wiley-Blackwell.

van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education, 85*, 175–189. https://doi.org/10.1016/j.tate.2019.06.010

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence (REL 2017–259)*. U.S. Department of Education, Institute of Education Sciences.

Konstantopoulos, S. (2008). The power of the test in three-level cluster randomised designs. *Journal of Research on Educational Effectiveness, 1*, 66–88.

Kozma, R. B. (2003). ICT and educational change: A global phenomenon. In R. B. Kozma (Ed.), *Technology, innovation, and educational change: A global perspective* (pp. 1–18). International Society for Technology in Education.

Lam, R. (2018). Testing, drilling and learning: What purpose does the Grade 3 Territory-wide System Assessment serve? *Asia Pacific Education Review, 19*(3), 363–374.

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2012). Structural equation modeling of mediation and moderation with contextual factors. In A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal*

studies (pp. 207–230). Routledge.
MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99–128.
McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
McMillan, J. H., Cohen, J., Abrams, L., Cauley, K., Pannozzo, G., & Hearn, J. (2010). *Understanding secondary teachers' formative assessment practices and their relationship to student motivation. (ERIC Document Reproduction Service No. ED507712)*.
McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement. *Practical Assessment, Research and Evaluation, 18*(2), 1–15.
Moss, C. M., & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders* (2nd ed.). ASCD.
Moss, C. M., Brookhart, S. M., & Long, B. A. (2013). Administrators' roles in helping teachers use formative assessment information. *Applied Measurement in Education, 26*(3), 205–218. https://doi.org/10.1080/08957347.2013.793186
Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4*(1), 56–78.
Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003
Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research. *Review of Educational Research, 74*, 525–556.
Rashid, R. A., & Jaidin, J. H. (2014). Exploring primary school teachers' conceptions of "assessment for learning". *International Education Studies, 7*(9), 69–83. https://doi.org/10.5539/ies.v7n9p69
Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam. *Curriculum Journal, 18*(1), 27–38.
Sach, E. (2015). An exploration of teachers' narratives: What are the facilitators and constraints which promote or inhibit 'good' formative assessment practices in schools? *Education 3-13, 43*(3), 322–335. https://doi.org/10.1080/03004279.2013.813956
Skerrett, A. (2010). 'There's going to be community. There's going to be knowledge': Designs for learning in a standardised age. *Teaching and Teacher Education, 26*(3), 648–655.
So, W. W. M., & Lee, T. T. H. (2011). Influence of teachers' perceptions of teaching and learning on the implementation of assessment for learning in inquiry study. *Assessment in Education: Principles, Policy & Practice, 18*(4), 417–432. https://doi.org/10.1080/0969594X.2011.577409
Song, E., & Koh, K. (2010). *Assessment for learning: Understanding teachers' beliefs and practices. [Paper presentation]. The 36th annual conference of the international association of educational assessment (IAEA) on "assessment for the future generations"*. August (Bangkok, Thailand).
Sussman, R., & Gifford, R. (2019). Causality in the theory of planned behavior. *Personality and Social Psychology Bulletin, 45*(6), 920–933.
Unal, A., & Unal, Z. (2019). An examination of K-12 teachers' assessment beliefs and practices in relation to years of teaching experience. *Georgia Educational Researcher, 16*(1). Article 2.
Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education, 37*(6), 747–760.
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.
Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Erlbaum.
Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934.
Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice, 22*(1), 140–160.
Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment, 6*(1), 133–158.
Yan, Z. (2014). School-based assessment in secondary schools. In C. Marsh, & J. C.-K. Lee (Eds.), *Asia's high performing education systems: The case of Hong Kong* (pp. 274–287). Routledge.
Yan, Z. (2021). Assessment-as-learning in classrooms: The challenges and professional development. *Journal of Education for Teaching, 47*(2), 293–295. https://doi.org/10.1080/02607476.2021.1885972
Yan, Z., & Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies In Educational Evaluation, 68*, 100985. https://doi.org/10.1016/j.stueduc.2021.100985
Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45*, 128–136.
Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). *A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. Assessment in Education: Principles, Policy & Practice*. https://doi.org/10.1080/0969594X.2021.1884042
Yan, Z., & Pastore, S. (2022). Assessing teachers' strategies in formative assessment: The Teacher Formative Assessment Practice Scale. *Journal of Psychoeducational Assessment*. https://doi.org/10.1177/07342829221075121.
Yan, Z., & Sin, K. F. (2014). Inclusive education: Teachers' intentions and behaviour analyzed from the viewpoint of the Theory of Planned Behaviour. *International Journal of Inclusive Education, 18*(1), 72–85. https://doi.org/10.1080/13603116.2012.757811
Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323–342.