

Relationships between EFL teachers' perceptions of consequential validity of formative assessment and data-driven decision-making self-efficacy and anxiety

Received 28 April 2023
Revised 11 June 2023
31 July 2023
Accepted 31 July 2023

Maryam Atai-Tabar, Gholamreza Zareian,
Seyyed Mohammad Reza Amirian and Seyyed Mohammad Reza Adel
Hakim Sabzevari University, Sabzevar, Iran

Abstract

Purpose – The purpose of this study was to ascertain the relationship between EFL teachers' perception of the intended and unintended consequences of formative assessment (FA) decisions and their sense of self-efficacy and anxiety toward data-driven decision-making (DDDM).

Design/methodology/approach – A correlational research design and correlational/regression analysis was utilized to conduct this study. In addition, a thematic analysis was conducted of participants' responses to the open-ended questions.

Findings – Descriptive statistics suggest that most EFL teachers perceived both intended and unintended consequential validity of their FA decisions with a moderate level of DDDM self-efficacy and a low level of anxiety. The results of the correlational analysis indicated a strong positive relationship indicating that those with higher teachers' perceptions of consequential validity of formative assessment (TPCVFA) scores tended to report higher DDDM efficacy. The results of the correlational analysis also indicated that a significant relationship did not exist between TPCVFA scores and DDDM anxiety. Finally, multiple regression analyses revealed that TPCVFA was a significant predictor of DDDM efficacy; however, TPCVFA was not a significant predictor of DDDM anxiety.

Originality/value – Data collected from 114 Iranian EFL teachers using the Data-Driven Decision-Making Efficacy and Anxiety Inventory (3D-MEA, Dunn *et al.* 2013a) and the TPCVFA questionnaire, which was self-developed and validated for the current study.

Keywords Formative assessment, Consequential validity, Data-driven decision-making, Intended and unintended consequences, Teachers' self-efficacy

Paper type Research paper

Introduction

The beneficial outcomes of formative assessment (FA), also known as assessment for learning, have been widely recognized by researchers and educators. Studies have shown the positive impact of FA that improves student motivation, engagement and self-regulation, and leads to a better understanding and retention of material. Additionally, FA can provide valuable information to teachers about student learning, which can be used to inform instructional decisions and provide targeted support (William, 2011; Dixon and Worrell, 2016; McCallum and Milner, 2021 and Remmi and Hashim, 2021).

Several studies particularly with regard to the nature of consequential validity have suggested that how different stakeholders can work toward improving the quality and impact of assessment in their classrooms and beyond (Gokturk Saglam and Tsagari, 2022;



Grocott, 2022; Slomp *et al.*, 2014 and Cowie, 2009). For classroom purposes, Popham (2003) and Winke (2011) also argued that the consequential validity of assessment refers to whether the use of assessment results, or the inferences based on a student's performance are valid. In this case, FA, validity is not the assessment instrument, rather the interpretation of the results that leads to valid educational decision-making (Van den Heuvel-Panhuizen and Becker, 2003). Teachers' use of assessment data to inform decisions has been recently defined as data-driven decision-making or DDDM (Gullo, 2013).

In line with the growing significance of assessment as an area of teaching and learning, teachers' self-efficacy is a crucial variable in assessment and making instructional decisions to amplify student learning. Despite the fact that the consequences of FA on teaching and learning, and the use of DDDM framework for informed instructional practice have been both theoretically and empirically discussed, few studies have been conducted to explore EFL teachers' self-efficacy for DDDM and their perceptions about the consequences of these decisions. The purpose of the current study was to propose a research agenda on the field of consequential validity of FA, and focus on the role of Iranian EFL teachers' thoughts and beliefs play in shaping educational decision making to close any teaching and learning gaps and help all students reach performance expectations.

Literature review

Teacher self-efficacy

Concerning teachers and classroom assessment practices, the role of self-efficacy or perception of competence in assessment (Bandura, 2000), has been emphasized. Individual differences in teacher self-efficacy appear essential to understanding and developing solutions to improve teacher classroom assessment practices (Bonner, 2016). Over the past 30 years, scholars have studied efficacy beliefs of teachers and have reported relationships between a teacher's sense of efficacy and a multitude of positive educational outcomes (Darling-Hammond *et al.*, 2013; Mirmojarabian and Rezvani, 2021). Regarding FA, it is supposed that teachers' self-efficacy can shape teacher understandings and experiences of assessment implementation and use (Eufemia, 2012). Thus, understanding and measurement of teacher-level factors such as self-efficacy around decision making is critical when it is considered as a learner-centered teaching tool (Dunn *et al.*, 2013).

Teacher self-efficacy for DDDM

Teachers' self-efficacy for DDDM refers to "teachers' beliefs in their abilities to organize and execute the necessary courses of action to successfully engage in classroom-level DDDM to enhance student performance" (Dunn *et al.*, 2013, p. 87). Teachers' self-efficacy in utilizing, analyzing and interpreting data to make informed decisions about their teaching, such as selecting instructional strategies, setting goals for student learning and monitoring student progress was imperative to understand the process of DDDM. Additionally, research has shown that teachers with high self-efficacy in DDDM are more likely to effectively use data to inform their instructional decisions, monitor student progress, plan lessons based on the data and make necessary changes to their teaching practices (Demchak and Sutter, 2019). This can promote a set of consequences corresponds to the intended uses of FA, such as fairness and equity, by ensuring that all students have access to high-quality instruction that is tailored to their individual needs (Ezzani, 2015).

Consequential validity of FA

Consequential validity was introduced by Messick (1989) and refers to the actual and potential consequences of assessment use. Assessment outcomes typically result in decisions

that carry intended and unintended positive and negative consequences (Kane, 2013; Slomp *et al.*, 2014). To advance the discussion, the concept of consequential validity of assessment was expanded by taking aspects such as the effects of assessment on teaching and learning and the consequences of how assessment information is used into consideration to reach a better, more in-depth and complete understanding of how a testing program functions (Gokturk Saglam and Tsagari, 2022). Earle (2021) argue that the validity in classroom assessment is about the accuracy of the interpretation and the use of assessment information. When thinking about validity, we focus on the inferences that are drawn from an assessment and the consequences of these inferences for those who have been assessed. Phakiti and Isaacs (2021) also contends that evaluating the validity and fairness of classroom assessments help language teachers improve their teaching and assessment practices. On the other hand, there are studies which show that if FA is misinterpreted or used for purposes that were not intended, the result may be poor decisions and problematic consequences (Gijbels and Dochy, 2006; Tsagari, 2009; Umer, 2015). A number of scholars have tried to examine the validity issue of FA, arguing that a major claim is the improvement of learning and teaching (Chalhoub-Deville and O'Sullivan, 2020).

Validity assurance of FA was studied by Cowie (2009) in terms of six desirable educational consequences: high expectations, inclusion, community engagement, coherence, learning to learn and a future focus. According to Cowie (2009), assessment is a tool that supports people to develop interpretations, make decisions and take action. Cowie's (2009) study describes in detail the complex contextualized nature of teachers' assessment decision making processes and provides a guide for the conceptualization of the positive consequences of FA.

In this research, FA to be consequentially valid will be seen through its impact on students' learning in terms of Cowie's (2009) ICCHLF principles. ICCHLF stands for inclusion, community engagement, coherence, high expectation, learning to learn and future focus. *Inclusion* refers to the importance of recognizing, valuing and engaging with students and ensuring that their learning and their assessment needs are met. *Community engagement* highlights the importance of teachers taking into account students' knowledge and practice to make their learning visible examining the gaps, fostering students 'active self-assessment and providing feedback to develop a shared understanding and enhance their learning.

Coherence provides students with clear learning pathway and progressions that allow them to meet learning goals and transition across activities. *High expectation* relates to the need for the curriculum, and teachers as the key decision makers, to empower all students to learn and achieve personal excellence. *Learning to learn* focuses on student learning, motivation and effort. And finally, *future focus* highlights the need for teacher assessment decision making and action to achieve a productive synergy between their short- and long-term goals for students learning and promotion of their teaching in particular.

In tandem with the above discussions in the field of FA, it is often hypothesized that the more confident teachers are about their assessment outcomes and their interpretations and uses, the more likely they will achieve the intended formative effects (Gu, 2021). However, exploring EFL teachers' perceptions of consequential validity of FA decisions, as well as their DDDM self-efficacy has been a relatively neglected topic in research. It is the intent of this study to add to this body of research by gathering information from EFL teachers about their self-efficacy in the process of DDDM and their perceptions about the validity of consequences of their educational FA decisions.

Methodology

Research design

The design of this study was descriptive correlational because it sought to provide information about particular variables (DDDM self-efficacy and teachers' perceptions of

consequential validity of formative assessment [TPCVFA]), within a single group sample (Iranian EFL teachers). In addition, the researcher also attempted to identify the nature of the relationship that exists between variables. A descriptive correlational design has the following characteristics: two variables are clearly identified and defined; data for each of two variables are collected; there is one group of participants; there is no intervention or treatment going on before, during or after data collection; data are collected at one general point in time; a correlation coefficient is calculated (Calmorin and Calmorin, 2007).

Participants

For this study, convenience sampling was adopted. One hundred and fourteen EFL teachers (males = 33% and females = 67%) were asked to fill out both the Data-Driven Decision-Making Efficacy and Anxiety Inventory (3D-MEA) and TPCVFA instrument. The teachers had at least five years of experience in teaching and assessment; 67% (N = 77) of the teachers completed the instruments online after they were identified by the researcher in online ELT groups. Paper version of the same instruments was also administered among teachers in seven institutions and universities in Iran. Thirty-three percent (N = 37) of the data from the teachers were obtained by administering the paper instruments. From the overall 114 EFL teachers who completed the instrument, 44.7% had been teaching in institutions, 32.1% at universities and 23.2% in both institutions and at universities at the time of data collection. They were from 20 to 56 years old (mean = 32.35). The teachers who participated in this study held a Bachelor's (25.4%), Master's (46.4%) and Ph.D. (28.2%). The majority had graduated from Teaching English as a Foreign Language/TEFL (79.3%); 7.2% had majored in English literature, 6.3% in translation and 1.8% in linguistics. The second part of survey including four open-ended questions was completed by 51 participants out of 114 (44.7%).

Instrumentation

Data collection was conducted via 3D-MEA (Dunn *et al.*, 2013). The 3D-MEA consists of 20 questions with a 5-point Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree*. The 3D-MEA was a validated and reliable instrument (Dunn *et al.*, 2013). Walker *et al.* (2018) also reported a high score internal consistency for all 5 scales in the 3D-MEA. The 3D-MEA 5 scales specifically denote questions of teacher efficacy in DDDM as follows: Efficacy for Data Identification and Access scale included 3 items that assessed a teacher's self-efficacy to identify, access and gather appropriate reports needed for formative DDDM. The Efficacy for Data Technology Use scale included 3 items that assessed a teacher's self-efficacy to utilize and navigate district and state level technology tools to access information for DDDM. The Efficacy for Data Analysis and Interpretation scale included 3 items that assessed a teacher's self-efficacy to analyze and interpret basic components of student performance data. The Efficacy for Application of Data to Instruction scale included 6 items that assessed a teacher's self-efficacy to connect and apply what was learned from data interpretation to instruction in order to improve student learning. The fifth scale, DDDM anxiety (Anxiety) included 5 items that assessed a teacher's sense of trepidation, tension and apprehension related to their ability to successfully engage in DDDM (Dunn *et al.*, 2013, p. 92).

The second instrument used was the TPCVFA questionnaire with 26 items which was developed by the current study, we drew on Cowie's (2009) conceptualization of consequential validity of FA. A pool of 43 Likert scale items in line with six consequential validity principles of Cowie (2009, ICCHLF) and comprehensive review of literature was developed by the researchers and checked for face validity by two EFL faculty members, and three faculty members specialized in assessment and evaluation. They classified 43 items in line with the Cowie's (2009) ICCHLF principles. Five-point Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree* was used for each selected item in the questionnaire to collect data from

respondents. The instrument was first piloted with three EFL teachers, and subsequently administered on a sample of 218 EFL teachers selected based on purposive sampling. Exploratory factor analysis procedures were conducted for the purpose of identifying the latent constructs underlying the data. Before EFA, the Kaiser–Meyer–Olkin (KMO) test and the Bartlett’s test of sphericity were conducted to evaluate the factorability or the sampling adequacy. The KMO value of 0.699 was greater than the recommended minimum values of 0.50 and the significance of Bartlett’s test of sphericity was less than 0.001, meaning that EFA can be applied to the obtained dataset (Field, 2009). To identify the number of factors to extract, parallel analysis which has shown to be the most accurate (Courtney, 2013; Pallant, 2011) was used. Parallel analysis for the current study was run in SPSS (v. 21) utilizing the rawpar.sps script developed by O’Connor (2000). Table 1 shows the results of the parallel analysis run on the 40 items, indicating that the retention of the three factors is appropriate.

Concerning parallel analysis, the size of the eigenvalues was compared with those obtained from a randomly generated data set of the same size to determine the number of factors and only those eigenvalues that exceeded the corresponding values from the random data set, were retained (Pallant, 2011). As shown in Table 1, the first three factors of the actual data have higher eigenvalues than the first three factors of the simulative data and as the other three factors, the eigenvalues of the simulative data are greater. The eigenvalue of the forth factor in the actual data is 0.674, whereas it is 1.098 in the simulative data. Because the eigenvalue of the simulative data of the fourth factor is higher than that of the actual data, the results of parallel analysis confirmed the presence of three components. It revealed only three components with eigenvalues exceeding the similar criterion values for a randomly generated data matrix of the same size. To further assess the possibility of a three-factor solution, a factor analysis with maximum likelihood extraction with avarimax rotation was then performed. Items with high loadings on more than one factor and items with small loadings on all factors were excluded from further analyses. This procedure resulted in a refined scale containing three subscales with 26 of the original 43 items. Each factor was then interpreted by examining item content align with Cowie’s (2009) principles and three factors were labeled as engagement, learning to learn and reflection on future-focused teaching. In line with Cowie (2009), the first factor is well matched with inclusion, and community engagement, the second factor with learning to learn and coherence, high expectation, the last factor with future focus principles. The factor loadings of all items were between 0.497 and 0.863. The three factors accounted for 54.39% of the total variance.

Confirmatory factor analysis (CFA) was then used to confirm the stability of the three-factor structure of 26-item TPCVFA questionnaire. As recommended by Gaskin and Lim (2016) and MacCallum *et al.* (2001), various fit indices such as the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA), were used to assess the adequacy of our model with a relatively small sample size. Using the new set of data (N = 107), the three-factor model for EFL teachers’ perceptions of consequential validity of FA decisions was tested using AMOS 23.

Factor	Eigen values of the actual data	Eigen values of the simulative data
1	2.173	1.459
2	2.059	1.300
3	1.811	1.187
4	0.674	1.098
5	0.553	1.035
6	0.515	0.964

Source(s): Authors’ own work

Table 1.
Eigen values of the actual data and the simulative data for TPCVFA

Table 2 shows that three of the indices (CMIN, RMSEA and SRMR) are within the acceptable range for model fit and confirm the absolute and parsimony fit of this model. Although CFI and TLI are lower than the acceptable value (0.84 and 0.89, respectively), TPCVFA can probably be considered as an acceptable model in terms of goodness of fit. The three-factor model also shows a nonsignificant correlation between the latent variables (0.51, 0.43 and 0.35, see Figure 1), which is indicative of discriminant validity. Moreover, results of reliability estimates of the overall scale and its dimensions showed that the Cronbach's α for the overall scale was 0.83 and for the dimension of engagement, learning to learn and reflection on future-focused teaching was 0.79, 0.88 and 0.91, respectively, suggesting satisfactory internal consistency of the developed items.

In addition, the questionnaire contains four open-ended questions so that the participants could comment on their perceived understanding of intended and unintended consequences of their decisions made from FAs. These questions developed after reviewing the literature and discussing the project with colleagues in the field. Adding open-ended questions to a quantitative self-report-based questionnaire can elicit much information and add significant value and depth to the results (Boeren, 2015). The questions were:

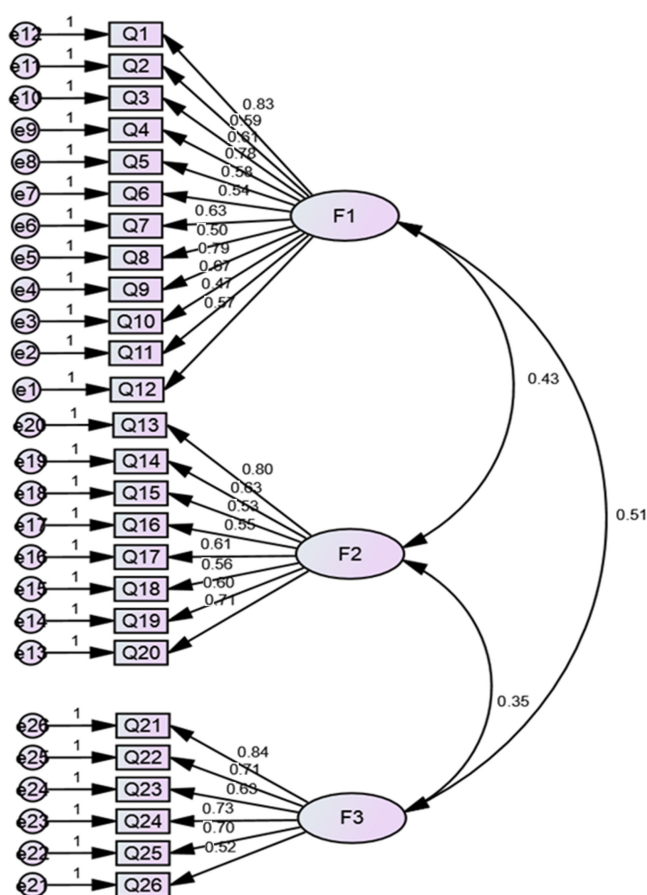
- (1) What types of instructional decisions do you make from FA practices?
- (2) Do these decisions lead to meaningful and appropriate outcomes for you and your students?
- (3) What are some unintended consequences of teachers' educational decisions made from FA results? And why do you think so?
- (4) Do the unintended consequences of teachers' educational decisions made from FA results impact students' learning? If so, how?

Data analysis

SPSS 19.0 packet program has been used for the analysis of data gained after the applications. Data were screened for missing, outlier and implausible values, and found to be normally distributed. Descriptive analysis including mean score was utilized on the data which was obtained after teachers' 3D-MEA and TPCVFA instruments administration. Thematic analysis was also adopted to extract the theme from the open-ended responses which was assumed to be informative in answering the research question of this study. A correlational analysis for three variables of TPCVFA and self-efficacy for DDDM, teacher anxiety for DDDM. Additionally, the researcher conducted a regression analysis in order to ascertain how much TPCVFA explained the teachers' sense of self-efficacy and anxiety for DDDM.

			Absolute fit indices			Comparative fit index		Parsimony fit index
	Df	p Value	CMIN/df	ChiSq(x2)	RMSEA	CFI	TLI	
Fitting dataset for three-factor model	296	0.00	1,182	1082.22	0.051	0.84	0.89	0.065
Acceptable threshold levels		p > 0.05	1 < CMIN < 5		close to 0.06 or below	> 0.95		< 0.08
Note(s): Modified TPCVFA after removal of experiences with factor loadings lower than 0.45 and modification indices								
Source(s): Authors' own work								

Table 2. Fit indices for three-factor CFA models of TPCVFA *



Source(s): Authors' own work

Figure 1.
Three-factor model of TPCVFA

Results

EFL teacher's self-efficacy for DDDM

The first part of the findings presents the descriptive analysis of the EFL teachers' responses to the five subscales of DDDM survey (N = 114). Each choice in this Likert-scale questionnaire was assigned a value ranging from 1 to 5. Five subscales of the 3D-MEA survey were analyzed and presented in Table 3. The mean total scale score is set to a value of 3.57 in a possible range of 1–5. The lowest DDDM self-efficacy score was for self-efficacy for data technology use (M = 2.97). The highest DDDM self-efficacy score was for self-efficacy for application of data to instruction (M = 4.39). DDDM anxiety scores were slightly low (M = 2.32), which revealed that EFL teachers disagreed with the belief that they were intimidated by the DDDM process.

EFL teacher's perceptions of consequential validity of their FA decisions

As another dimension of the study, the respondents were administered a self-developed survey (N = 114) with three subscales. Mostly, the respondents agreed with the statements

included in the TPCVFA instrument were supposed to elicit the participants' perceptions with regard to the consequential validity of their FA decisions. The average mean of their responses was 3 and above, which indicates interesting patterns.

The statements included in Table 4 were supposed to elicit the participants' opinion with regard to community engagement as a valid consequence of their FA decisions. The respondents have mostly agreed with the statements. The overall mean score for the first subscale of the TPCVFA questionnaire equaled 3.76. Therefore, on one hand there were two items with the highest mean scores 5 ($M = 4.24$), 8 ($M = 4.21$), in which they asserted that they (a) explicitly encourage further learning and (b) select interventions for gaps in student understanding, respectively. On the other hand, the lowest mean scores belonged to items 10 ($M = 2.61$), through which the teachers remarked that they do not modify or adapt their syllabus and/or lesson plans based on students' needs.

Table 5 indicated the EFL teachers' perceptions of the consequences of their FA decisions regarding learning to learn. A clear majority of the respondents for items 18 ($M = 4.78$), 17 ($M = 4.65$) and 15 ($M = 4.28$), reported that their FA decisions cause them to (a) foster student motivation and effort, (b) help students identify gaps in their learning and (c) guide student judgment of what is important to learn respectively. In addition, they stated difficulty in creating conditions for students' self-reflection and review of ideas (item 19, $M = 3.08$).

In Table 6, almost half of the respondents were of the view that engaging in reflection to make corrective moves (Park and Datnow, 2017) in teaching and assessment was important as they mostly agreed with the statements. Concerning the importance of the consequence of FA decisions as reflection in items 22 ($M = 4.43$), 25 ($M = 4.36$), 21 ($M = 4.12$) and 26

Table 3.
Descriptive statistics of
EFL teachers' self-
efficacy for
TPCVFA scale

Subscales	N	Mean	Std. Deviation	Std. Error mean
Data identification and access	114	4.07	0.82	0.07
Data technology use	114	2.97	1.00	0.10
Data analysis and interpretation	114	4.12	0.74	0.07
Application of data to instruction	114	4.39	0.68	0.06
Data-driven decision making anxiety	114	2.32	1.24	0.09
Total	114	3.57	0.97	0.08

Source(s): Authors' own work

Table 4.
TPCVFA regarding
community
engagement

Items	Statements for community engagement consequences	M	St.D
	1. Identifying gaps in knowledge and understanding	4.18	0.73
	2. Learning needs and preconceptions	4.12	0.72
	3. Projecting students forwards or re-teaching	3.99	0.46
	4. Causes of both strengths and shortcomings	3.96	0.68
	5. Explicitly encouraging further learning	4.24	0.78
	6. Descriptive feedback on performance/progress	3.66	0.68
	7. Interventions to drive instruction	4.13	0.72
	8. Interventions for gaps in understanding	4.21	0.72
	9. Support and encouragement	3.37	0.57
	10. Modifying and adapting syllabus and/or lesson plans	2.61	0.58
	11. Personal excellence	3.01	0.80
	12. Further leaning experiences	3.44	0.58
	Total	3.76	0.64

Source(s): Authors' own work

($M = 4.11$), the teachers affirmed that they intended to use formative data for (a) determining effective teaching and assessment strategies, (b) identifying gaps in their instructional curriculum, (c) making appropriate improvements and (d) evaluating the effectiveness of their instruction.

EFL teachers' perceptions of formative assessment

The relationship between EFL teachers' self-efficacy for DDDM and their perceptions about consequential validity of FA decisions

Correlational analyses were conducted to determine the relationship between the TPCVFA scores and DDDM efficacy and DDDM anxiety scores (see Table 7). We also used the guideline of correlation value based on Evans (1996). The results from the correlational analysis demonstrated a strong positive relationship, indicating that those teachers with higher TPCVFA scores tended to report higher DDDM efficacy ($r = 0.611, p \leq 0.01$). The results of the correlational analysis also demonstrated that there was no significant

Items	Learning to learn consequences	M	St.D
	13. Needs-based participation	3.11	0.61
	14. Planning for learning	3.22	0.89
	15. Judgment	4.28	0.65
	16. Confidence	3.69	0.76
	17. Identifying gaps by students	4.65	0.65
	18. Motivation and effort	4.78	0.69
	19. Self-reflection and review of ideas	3.08	0.73
	20. Monitoring achievement growth/progress over time	4.05	0.52
	Total	3.85	0.71

Source(s): Authors' own work

Table 5.
TPCVFA Regarding Learning to learn

	Reflection on future-focused teaching	M	St.D
	21. Reflection on assessment practice	4.12	0.641
	22. Effective teaching and assessment strategies	4.43	0.74
	23. Continuity and coherence	3.01	0.52
	24. Instruction improvement	3.69	0.71
	25. Identifying instructional gaps	4.36	0.63
	26. Evaluating instruction effectiveness	4.11	0.69
	Total	3.95	0.68

Source(s): Authors' own work

Table 6.
TPCVFA regarding reflection on future-focused teaching

	Consequential validity of FA decisions	Self-efficacy for DDDM	Anxiety
Consequential validity of FA decisions	–	0.611 **	–0.026
Self-efficacy for DDDM	0.611 **	–	–0.105
Anxiety	–0.026	–0.105	–

Note(s): **Correlation is significant at the 0.01 level (2-tailed)

Source(s): Authors' own work

Table 7.
Correlation analysis of consequential validity of FA decisions, DDDM self-efficacy and DDDM anxiety

relationship between the TPCVFA scores and DDDM anxiety ($r = -0.026, p > 0.05$). The closeness of the r value to zero indicates that almost no relationship was present between the variables.

Following the correlational analyses, multiple regression analyses were conducted using the variables of TPCVFA, DDDM efficacy to address research question three. The multiple regression analysis utilized the DDDM efficacy score as the criterion variable, the TPCVFA score as the predictor variable to see what percentage of variability in teacher sense of DDDM efficacy can be explained by TPCVFA. The standardized regression coefficients were 0.610 for TPCVFA (Table 8). The coefficient of determination, R^2 , of 0.372 was significant ($p < 0.05$). The adjusted R^2 of 0.372 that compensates for the positive bias indicates that approximately 37% of the variability of DDDM efficacy can be explained by TPCVFA. However, TPCVFA was not a significant predictor of DDDM anxiety.

A more in-depth examination: comments from EFL teachers

Thematic analysis of the open-ended questions revealed the repeated themes emerging from teachers' comments on the types of consequences of their FA decisions and then categorized into five key issues as follows: instructional placement, diagnosing student needs, monitoring their progress and providing appropriate instructional interventions, measuring student's growth. They tended to make valid decisions or take appropriate actions for learning and teaching improvement. They commented that they could use assessment results to understand the learning and teaching practice better and address the requirements as a whole. They also mentioned that they can use the assessment results to reflect on student progress and take assessment results into the design of future instruction. However, their concerns about external factors (i.e. an increased focus on the accountability requirements, lack of time and large class size), which are out of their control may have led to lower levels of their DDDM self-efficacy in their actual FA practice. A summary of the results of the teachers' comments are following in Table 9.

Table 8.
Regression analysis of
DDD self-efficacy on
EFL teachers'
perceptions of FA
consequential validity

Model	N	R	R^2	Adjusted R^2	Sig
	114	0.611	0.373	0.370	0.048
Source(s): Authors' own work					

Table 9.
Teachers' assessment
decisions, intended and
unintended
consequences

	Educational decisions	Intended consequences	Unintended consequences
Repeated themes emerged from EFL teachers' comments	Placement Need analysis Diagnostic Instruction Progress monitoring of the whole class	Examining the gaps Appropriate instruction Empowering all students to write better Supportive learning Personalized learning Reflective practice	Emotional problems anxiety Lack of motivation Less creativity Less flexibility Fail to progress Teaching to the test Accountability Improvement of test score and graduation outcome rather than mastering of content
Source(s): Authors' own work			

Discussion and conclusion

Many EFL teachers struggle with FA implementation and DDDM practices to make ongoing decisions to optimize instruction ([Hamilton and Reeves, 2022](#)). Our study was supposed to connect EFL teachers DDDM self-efficacy with the FA consequences they believe to achieve to impact subsequent engagement. That is, if teachers believe that FA played its role as a teaching tool and influenced what and how the students learned, they may experience performance accomplishments and increase their self-efficacy related to DDDM. Therefore, the purpose of this study was to provide more insight into the interplay between EFL teachers' perceptions of the consequential validity (intended and unintended consequences) of their FA decisions and DDDM self-efficacy. The first research question pertained to the Iranian EFL teachers' perceptions of the consequential validity of their FA decisions suggesting [Cowie's \(2009\)](#) model for understanding the elements of the FA consequential validity, within the context, through which they impart their effects and the final outcomes obtained from FA decisions.

This study was also an attempt to explore Iranian EFL teachers' self-efficacy beliefs and anxiety in the area of DDDM. The results showed that Iranian EFL teachers had a moderate level of DDDM self-efficacy and a low level of anxiety. While DDDM self-efficacy has been playing an increasing role in contemporary teaching, enhancing students' performance and supporting educational matters ([Marsh *et al.*, 2006](#); [Datnow and Hubbard, 2016](#) and [Mandinach and Gummer, 2016b](#)), there has been no published empirical research having investigated DDDM self-efficacy and anxiety particularly in Iranian higher education context to provide an optimal arena in which to address EFL teachers' DDDM efficacy and DDDM concerns.

The findings implied that the teachers' perceptions are as one of the important predictors of DDDM self-efficacy, especially perceptions of their own FA practices. Just as teachers' beliefs and perceptions associated with any interventions worthy of their time and effort influence their DDDM behavior in the classroom ([Reeves and Chiang, 2018](#)), it was assumed that teachers' perceptions about what FA on its own may yield as intended consequences would also be of value to DDDM self-efficacy and anxiety. Once this accomplished, teacher's DDDM self-efficacy will increase to trigger the intended consequences required. Therefore, the EFL teachers' perceptions of the validity of their FA decisions measured in the TPCVFA are critical to developing teachers DDDM self-efficacy and anxiety. Apparently, this self-efficacy for DDDM was not grounded in a complete understanding of the intended consequences of FA decisions to improve instruction, increase program effectiveness and guide student learning. It seems that teachers' lower level of self-efficacy due to the external matters is a concern. They believe that the external factors (i.e. time, class size, emotional filters and accountability system) are out of their control and prevent them from making instructional decisions that mediate the intended consequences of assessments in practice. They comment pressures of external factors have the possibility of conflicting with the reform efforts that try to provide assessment training to teachers to ensure that instructional time is spent delivering the most effective, research-based assessment strategies. As [Mellati and Khademi \(2018\)](#) mentioned, positive attitudes toward the significance of conducting instructional assessment, FA, (as a source of teachers' self-efficacy) inform EFL teachers about the importance of advocating for more effective instructional assessment. Therefore, it will be fruitful to provide professional development resources for EFL teachers beyond training stage to monitor the impact of assessment reform on student performance, which in turn impact teachers' beliefs or attitudes positively in spite of exposing to the pitfalls of uncontrolled conditions.

The findings of previous teacher self-efficacy studies in the area of assessment studies implied the notion that teachers' higher levels of self-efficacy scores for DDDM was expected because of mediating effect of attitude and teaching experience and assessment knowledge

and practice (Buldur and Tatar, 2011; Dunn *et al.*, 2013; Eslami and Fatahi, 2008; Mirmojarabian and Rezvani, 2021 and Oo *et al.*, 2021). From this standpoint, Buldur and Tatar (2011) strongly emphasized the importance of training both in-service and pre-service teachers on the design and use of assessment strategies and explained that without proper training and education, teachers might not feel efficacious enough to persist in implementing DDDM.

The research described here clearly has some limitations. First, the total number of the participants in this study was 114 EFL teachers in Iran. Due to limited participant descriptions, as well as a limited geography, the study cannot be said to have high generalizability. Thus, the findings of this study were treated with caution. Future research focusing on a greater number of participants should be considered. This study suffers from the problems inherent in using questionnaires. As any study utilizing survey data, the present study is not exempt from problems associated with the issue of central tendency bias (participants avoiding extreme response categories), acquiescence bias (participants agreeing with certain statements to please the researcher) and social desirability (participants portraying themselves in a more socially favorable light rather than being honest) (Gu, 2016; Subedi, 2016; Van de Vijver and Leung, 2000). So, the second limitation was related to the exclusion of qualitative methods such as focus group discussions and individual interviews for data collection. To decrease the influence of the biases, mixed-methods studies are suggested to portray a fine-grained picture of teachers' self-efficacy to use assessment data for intended consequences and decide on appropriate next steps. Finally, as a new and promising instrument the TPCVFA underscores the need of further construct validation work. In this regard, we hope other researchers use the TPCVFA and report the usefulness and performance of this instrument to measure the extent to which teachers believe that they purposefully pursue the desired effect on students' learning and achievement.

References

- Bandura, A. (2000), "Regulation of cognitive processes through perceived self-efficacy", *Psychology of Education*, Vol. 2, pp. 365-380.
- Boeren, E. (2015), "Surveys as tools to measure qualitative and quantitative data", in *Handbook of Research on Scholarly Publishing and Research Methods*, IGI Global, pp. 415-434.
- Bonner, S.M. (2016), "Teachers' perceptions about assessment: competing narratives", in Brown, G.T.L. and Harris, L.R. (Eds), *Handbook of Human and Social Conditions in Assessment*, Routledge, pp. 37-55.
- Buldur, S. and Tatar, N. (2011), "Development of self-efficacy towards using alternative assessment scale", *Asia Pacific Education Review*, Vol. 12, pp. 485-495.
- Calmorin, L.P. and Calmorin, M.A. (2007), *Research Methods and Thesis Writing*, Rex Bookstore, Manila.
- Chalhoub-Deville, M. and O'Sullivan, B. (2020), *Validity: Theoretical Development and Integrated Arguments*, British Council Monographs, Sheffield.
- Courtney, M.G. (2013), "Determining the number of factors to retain in EFA: using the SPSS R-menu v2.0 to make more judicious estimations", *Practical Assessment, Research, and Evaluation*, Vol. 18 No. 1, pp. 1-14.
- Cowie, B. (2009), "Teacher FA decision making: a consideration of principles and consequences", *Assessment Matters*, Vol. 1, pp. 47-64.
- Darling-Hammond, L., Newton, S.P. and Wei, R.C. (2013), "Developing and assessing beginning teacher effectiveness: the potential of performance assessments", *Educational Assessment, Evaluation and Accountability*, Vol. 25 No. 3, pp. 179-204.
- Datnow, A. and Hubbard, L. (2016), "Teacher capacity for and beliefs about data-driven decision making: a literature review of international research", *Journal of Educational Change*, Vol. 17, pp. 7-28.

-
- Demchak, M. and Sutter, C. (2019), "Teachers' perception of use and actual use of a data-based decision-making process", *Education and Training in Autism and Developmental Disabilities*, Vol. 54 No. 2, pp. 175-185.
- Dixson, D.D. and Worrell, F.C. (2016), "Formative and summative assessment in the classroom", *Theory Into Practice*, Vol. 55 No. 2, pp. 153-159.
- Dunn, K.E., Airola, D.T., Lo, W.J. and Garrison, M. (2013), "What teachers think about what they can do with data: development and validation of the data driven decision-making efficacy and anxiety inventory", *Contemporary Educational Psychology*, Vol. 38 No. 1, pp. 87-98.
- Earle, S. (2021), "Formative decision-making in response to primary science classroom assessment: what to do next?", *Frontiers in Education*, Vol. 5, 584200.
- Eslami, Z.R. and Fatahi, A. (2008), "Teachers' sense of self-efficacy, English proficiency, and instructional strategies: a study of nonnative EFL teachers in Iran", *TESL-EJ*, Vol. 11 No. 4, pp. 1-19.
- Eufemia, F. (2012), *The Relationship between FA and Teachers' Self-Efficacy*, [Unpublished doctoral dissertation], University of Florida, Gainesville, FL.
- Evans, J.D. (1996), *Straightforward Statistics for the Behavioral Sciences*, Thomson Brooks/Cole Publishing.
- Ezzani, M. (2015), "Coherent district reform: a case study of two California school districts", *Cogent Education*, Vol. 2 No. 1, pp. 1-20.
- Field, A. (2009), *Discovering Statistics Using SPSS: Introducing Statistical Method*, 3rd ed., Sage, Thousand Oaks.
- Gaskin, J. and Lim, J. (2016), "Master validity tool, AMOS plugin", *Gaskination's StatWiki*, Vol. 37 No. 3, pp. 814-822.
- Gijbels, D. and Dochy, F. (2006), "Students' assessment preferences and approaches to learning: can FA make a difference?", *Educational Studies*, Vol. 32 No. 4, pp. 399-409.
- Gokturk Saglam, A.L. and T sagari, D. (2022), "Evaluating perceptions towards the consequential validity of integrated language proficiency assessment", *Languages*, Vol. 7 No. 65, pp. 2-18.
- Grocott, C. (2022), "Intended or unintended consequences? Assessing the ways in which the National Tests of English are implemented and acted on in Norwegian schools", *Nordic Journal of Language Teaching and Learning*, Vol. 10 No. 1, pp. 25-51.
- Gu, P.Y. (2016), "Questionnaires in language teaching research", *Language Teaching Research*, Vol. 20 No. 5, pp. 567-570.
- Gu, P.Y. (2021), "An argument-based framework for validating formative assessment in the classroom", *Frontiers in Education*, Vol. 6, pp. 1-10.
- Gullo, D.F. (2013), "Improving instructional practices, policies, and student outcomes for early childhood language and literacy through data-driven decision making", *Early Childhood Education Journal*, Vol. 41, pp. 413-421.
- Hamilton, V.M. and Reeves, T.D. (2022), "Relationships between course taking and teacher self-efficacy and anxiety for data-driven decision making", *The Teacher Educator*, Vol. 57 No. 2, pp. 136-154.
- Kane, M.T. (2013), "Validating the interpretations and uses of test scores", *Journal of Educational Measurement*, Vol. 50 No. 1, pp. 1-73.
- MacCallum, R.C., Widaman, K.F., Preacher, K.J. and Hong, S. (2001), "Sample size in factor analysis: the role of model error", *Multivariate Behavioral Research*, Vol. 36 No. 4, pp. 611-637.
- Mandinach, E.B. and Gummer, E.S. (2016), "What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions?", *Teaching and Teacher Education*, Vol. 60, pp. 366-376.
- Marsh, J.A., Pane, J.F. and Hamilton, L.S. (2006), *Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research*, RAND Corporation, Santa Monica, CA.

-
- McCallum, S. and Milner, M.M. (2021), "The effectiveness of formative assessment: student views and staff reflections", *Assessment and Evaluation in Higher Education*, Vol. 46 No. 1, pp. 1-16.
- Mellati, M. and Khademi, M. (2018), "Exploring teachers' assessment literacy: impact on learners' writing achievements and implications for teacher development", *Australian Journal of Teacher Education*, Vol. 43 No. 6, pp. 1-12.
- Messick, S. (1989), "Meaning and values in test validation: the science and ethics of assessment", *Educational Researcher*, Vol. 18 No. 2, pp. 5-11.
- Mirmojarabian, M.S. and Rezvani, E. (2021), "EFL teachers' self-efficacy beliefs and their (non)-communicative instructional practices", *TEFLIN Journal*, Vol. 32 No. 2, pp. 267-294.
- Oo, C.Z., Alonzo, D. and Davison, C. (2021), "Pre-service teachers' decision-making and classroom assessment practices", *Frontiers in Education*, Vol. 6, pp. 1-12.
- O'Connor, B.P. (2000), "SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test", *Behavior Research Methods, Instruments and Computers*, Vol. 32 No. 3, pp. 396-402.
- Pallant, J. (2011), *Survival Manual. A Step by Step Guide to Data Analysis Using SPSS*, 4th ed., Allen & Unwin, Crows Nest, NSW.
- Park, V. and Datnow, A. (2017), "Ability grouping and differentiated instruction in an era of data-driven decision making", *American Journal of Education*, Vol. 123 No. 2, pp. 281-306.
- Phakiti, A. and Isaacs, T. (2021), "Classroom assessment and validity: psychometric and edumetric approaches", *European Journal of Applied Linguistics and TEFL*, Vol. 10 No. 1, pp. 3-24.
- Popham, W.J. (2003), *Test Better, Teach Better: the Instructional Role of Assessment*, Association for Supervision and Curriculum Development, Alexandria, VA.
- Reeves, T.D. and Chiang, J.L. (2018), "Online interventions to promote teacher data-driven decision making: optimizing design to maximize impact", *Studies in Educational Evaluation*, Vol. 59, pp. 256-269.
- Remmi, F. and Hashim, H. (2021), "Primary school teachers' usage and perception of online FA tools in language assessment", *International Journal of Academic Research in Progressive Education and Development*, Vol. 10 No. 1, pp. 290-303.
- Slomp, D.H., Corrigan, J.A. and Sugimoto, T. (2014), "A framework for using consequential validity evidence in evaluating large-scale writing assessments: a Canadian study", *Research in the Teaching of English*, Vol. 48 No. 3, pp. 276-302.
- Subedi, B.P. (2016), "Using likert type data in social science research: confusion, issues and challenges", *International Journal of Contemporary Applied Sciences*, Vol. 3 No. 2, pp. 36-49.
- Tsagari, D. (2009), *The Complexity of Test Washback: an Empirical Study*, Peter Lang GmbH, Frankfurt am Main.
- Umer, M. (2015), "FA and consequential validity: a practice yet to be effectively implemented in Saudi higher education", *Pakistan Journal of Education*, Vol. 32 No. 2, pp. 345-356.
- Van de Vijver, F.J.R. and Leung, K. (2000), "Methodological issues in psychological research on culture", *Journal of Cross-Cultural Psychology*, Vol. 31, pp. 33-51.
- Van den Heuvel-Panhuizen, M. and Becker, J. (2003), "Towards a didactic model for assessment design in mathematics education", in Bishop, A.J., Clements, M.A., Keitel, C., Kilpatrick, J. and Leung, F.K.S. (Eds), *Second International Handbook of Mathematics Education*, Springer, pp. 689-716.
- Walker, D.A., Reeves, T.D. and Smith, T.J. (2018), "Confirmation of the data driven decision-making efficacy and anxiety inventory's score factor structure among teachers", *Journal of Psychoeducational Assessment*, Vol. 36 No. 5, pp. 477-491.
- Wiliam, D. (2011), "What is assessment for learning?", *Studies in Educational Evaluation*, Vol. 37 No. 1, pp. 3-14.
- Winke, P. (2011), "Evaluating the validity of a high-stakes ESL test: why teachers' perceptions matter", *TESOL Quarterly*, Vol. 45 No. 4, pp. 628-660.

About the authors

Dr Maryam Atai-Tabar has Ph.D. in applied linguistics. She does research in dynamic assessment, teachers assessment literacy and formative assessment. she is interested in English for academic purposes, English for research publication purposes and reflective teacher education. Maryam Atai-Tabar is the corresponding author and can be contacted at: maria.ataii2018@gmail.com

Dr Gholamreza Zareian is currently Associate Professor of TEFL at Hakim Sabzevari University. His main research interests are materials development, ESP and individual differences. He has published some articles in both local and international refereed journals and some textbooks for university students.

Dr Seyyed Mohammad Reza Amirian is Associate Professor of TEFL at Hakim Sabzevari University, Iran. He has taught graduate and postgraduate courses such as advanced testing, research methodology and language skills. His research interests include testing and assessment, teacher education, psycholinguistics, vocabulary and DIF.

Dr Seyyed Mohammad Reza Adel has Ph.D. in TEFL from Ferdowsi University in Mashhad, Iran. He is currently Associate Professor at Hakim Sabzevari University in Iran. His main area of research interest includes socio-applied linguistics. He has presented and published papers at many national and international conferences and leading journals such as Teaching in Higher Education, Qualitative Research Journal, Journal of Intercultural Communication Research and Higher Education. He has taught graduate and postgraduate courses for more than a decade. He has also supervised some M.A and Ph.D. theses to completion.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com