Mohammadi Darabad *et al.*
Language Testing in Asia (2023) 13:13
https://doi.org/10.1186/s40468-023-00218-4

Language Testing in Asia

RESEARCH Open Access

On the construct and perceptual validity measures of L1-based vs. L2-based elicitation as a measure of L2 classroom performance assessment

Ali Mohammadi Darabad[1], Gholam-Reza Abbasian[2*], Bahram Mowlaie[1] and Ali-Asghar Rostami Abusaeedi[3]

*Correspondence:
gabbasian@gmail.com

[1] Faculty of Persian Literature and Foreign Languages, Islamic Azad University, South Tehran Branch, Tehran, Iran
[2] Imam Ali University, Tehran, Iran
[3] English Department, Faculty of Literature and Humanities, Shahid Bahonar University of Kerman, Kerman, Iran

Abstract

This study aimed at investigating the English as a foreign language (EFL) learners' perceptions of L1-based and L2-based elicitations in the English classroom employing an explanatory sequential mixed-method design. Ninety-seven Iranian intermediate EFL learners of English have been selected from Islamic Azad University (Science and Research Branch) in Tehran Province using a convenient sampling method. Of these, in the qualitative phase, 15 individuals were selected through a convenience sampling method as the focus group ($N=15$; $n=8$ for the L1 group and $n=7$ for the L2 group). In the quantitative phase, 90 intermediate EFL learners were selected. The selected participants' L2 performances were assessed through L1-based and L2-based elicitation techniques. They completed two validated researcher-made questionnaires to capture their perceptions of the elicitation techniques. Accordingly, five separate exploratory factor analyses were run to investigate the underlying constructs of the five components of the L1-based and L2-based perception questionnaires, the results of which showed that the correlation matrices were not singular and there were perfect correlations among all variables of L1-based and L2-based perception questionnaires. The findings show that the majority of respondents prefer to use their L2 in speaking classes and believe that L2 should be the prior language in general speaking classes. More than two-thirds of the respondents prefer to use L2 when they want to communicate with each other inside and outside of the classroom and prefer to use L2 in doing their assignments or performing orally in class. More than half of the respondents prefer to use L2 in assessment sessions.

Keywords: EFL learners, L1-and L2-based elicitations, Perceptions, Performance assessment

Introduction

The use of the student's first language (L1) in the EFL classroom has long been debated on the grounds of contradictory findings. The existing contradictory findings revolve around the context-related nature of language learning. Some authors advocate the use of L1 in English classes (e.g., Auerbach, 1993; Schweers, 1999), while others negate

the role of L1 use. Some other scholars (e.g., Ellis, 1984) hold a moderate position that too much L1 use could be a barrier to both language learning and learners against L2 exposure.

In the Iranian EFL setting, the only experience of using the target language happens in the classroom. Consequently, it is common for EFL teachers to use the students' mother tongue as a tool to convey the message as a means of interaction. Although many researchers (e.g., Chiou, 2014; Liu & Zeng, 2015; Shin et al. 2019) believe that teaching through target L2 gives better results, the issue requires further research in general and in the Iranian EFL setting in particular.

Different aspects of L1 use have been investigated through many empirical studies in EFL education. For example, Mohammadi Darabad et al. (2021) investigated the L1-based elicitation technique concerning L2 performance assessment, seeking the validity measures of transfer. Their findings showed that the L1-based elicitation technique was a valid measure of L2 performance assessment. However, validity is preferred to be approached from a unitary perspective (Messick, 1989). In this sense, a validity claim is not supported solely by a single source of evidence. Considering construct validity as the central and pivotal constituent in validation studies, five sources of evidence are still required to make valid use of the interpretation and use of the score of a testing instrument. These sources include evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, empirical or criterion-related validity, and consequences. Based on this, the stakeholders' judgment can also be taken as a source of inference. This is one of the many points left so far intact in the literature in general and in the Iranian EFL context in particular. To spot this gap, this study aims at investigating the perceptual validity of the L1-based elicitation technique as a teaching device in an L2 classroom from the learners' perspectives. Accordingly, an attempt has been made to explore the students' perceptions of L1-based and L2-based elicitations as the construct of the questionnaires using various statistical techniques, including factor analysis. The validated questionnaires then were conducted to examine the participants' perceptions of the L1-based and L2-based elicitations concerning their L2 performance assessment.

Literature review

The discussion about whether the first language should be involved or not in language classrooms has been an argumentative topic for so long. This discussion revolves around the L1 use (Cook, 2008; Mohebbi & Alavi, 2014), the role of learners' first language in the foreign language classroom (Al Sharaeai, 2012; Littlewood & Shufang, 2022; Rivers, 2011; Tsagari & Diakou, 2015; Turnbull, 1999), functions of L1 in L2 learning (Molway et al., 2022), language transfer and skills, i.e., speaking, listening, reading, and writing (Perkins & Zhang, 2022), and language teachers' and learners' perceptions of L1 use in an L2 classroom context, elicitation techniques, assessment mechanisms, and validity issues concerning the assessment mechanism and perceptions. It seems that these contentious subjects will be at the center of the educational agenda in the future. Regarding the use of L1 and focusing on various areas of language features and skills, i.e., morpho-syntax, vocabulary, phonology, reading, and writing, many researchers (e.g., Eckman, 2014; Lardiere, 2014; McManus, 2022; Perkins & Jiang, 2019; Perkins & Zhang, 2022;

Polio, 2014; Ringbom & Jarvis, 2009) concluded that encouraging learners to attend to similarities and differences between L1 and L2 is an important instructional strategy for improving L2 abilities. They also concluded that learners depend on perceiving L1-L2 similarities between individual items and their functional equivalences between two underlying grammatical systems. Therefore, similarities and differences between L1 and L2 play an important role.

Emphasizing the importance of collecting the stakeholders' perspectives for validation (e.g., Bachman & Palmer, 2010; Ferrando, 2013; Gokturk Saglam & Tsagari, 2022; Im et al., 2019; Macaro & Lee, 2013; Nguyen & Habok, 2021; Shohamy, 2001; Yao, 2011), some attitudinal studies, though sporadically, have been reviewed. Some researchers (e.g., Afzal, 2012; Alshammari, 2011; Khati, 2011; Saito & Ebsworth, 2004; Sharma, 2006) concluded that many teachers and learners prefer to employ L1 to explain new vocabulary, concepts, and grammar rules, to give instructions for activities, to understand the subject, and to communicate with the teacher or other students. In some other local studies (e.g., Mahmoudi & Amirkhiz, 2011; Nazary, 2008), students believed that the dominant language in English classrooms should be English rather than the students' L1, and the students were reluctant to use their L1 in English class.

In another study, Molway et al. (2022) focused on teachers' reported practices regarding the amounts and functions of L1 and L2 use in the L2 classroom and explored some of the many possible factors shaping those practices, including experiences during preservice training, number of years in service and national context in terms of language education policy, and the social value of the L2s. Their main finding is that teachers' reported practices vary significantly by location (whether they were teaching in Spain or in England). Teachers of English in Spain report more frequent use of the L2 across all classroom language functions investigated (i.e., grammar teaching, giving details about tests and exams, and teaching cultural content). Despite the existing body of research regarding the use of L1 for grammar teaching in both L1 Spanish and L1 English contexts, over 50% of teachers based in Spain reported a predominant use of L2 for this function. Regarding details about tests and exams, the teachers in England made use of L1 to ensure student comprehension of critical aspects of examination techniques. The teachers also use L1 for administrative matters to save valuable classroom time for more meaningful learning opportunities. Finally, regarding teaching cultural content, teachers in England reported very high levels of L1 use for this purpose.

Concerning the elicitation and assessment mechanisms, Gass (2018) believes that elicitation methods and the types of eliciting data are selected by the researchers to understand how the languages are learned. Two common elicitation tasks, namely judgments and elicited imitation, were reviewed by Gass (2018). The effectiveness of four types of judgment tasks, including magnitude estimation, grammaticality judgments, truth-value judgments, and preference judgments, was investigated by Plonsky et al. (2019). Their findings supported the effectiveness of these elicitation types on L2 assessment. Investigating the ways of language proficiency measurement, Wu and Ortega (2013) and Gaillard and Tremblay (2016) emphasized the use of elicited imitation as a worthy measure of general proficiency. This claim was also advocated by Yan et al.'s (2016) meta-analysis study in which 21 studies were analyzed. According to the obtained results, elicited imitation was a discrimination factor across proficiency levels.

Concerning validity issues, Stansfield and Kenyon (1992) examined the concurrent validity of semi-direct and direct tests in a number of languages, the results of which demonstrated high correlations between the two types of tests. The use of semi-direct tests was recommended as a valid and practical substitute for direct tests. Wigglesworth and O'Loughlin (1993) investigated the comparability of two versions of an oral interaction test, i.e., a tape-based (semi-direct) version, and a live interview (direct) version. They showed that the two versions were highly comparable.

Using quantitative and qualitative procedures, Shohamy (1994) explored the validity of semi-direct versus direct tests. The correlational analyses revealed high concurrent validity of the two tests (Shohamy & Stansfield, 1991; Shohamy et al., 1989); however, the tests differed in a number of aspects. Qualitative analyses specified that the differences were in the topics, number of functions employed in the elicitation tasks, and the communicative strategies, i.e., more paraphrasing and self-correction on the semi-direct test and more shifts to L1 resources on the direct test.

These studies and some more have shown that there has been a judicious quantity of research on L1 use in English classrooms, and a majority of them have addressed the teachers' ideas and attitudes toward the issue; however, less research has focused on the student's perceptions of the L1 use in these settings. The issue has been investigated more in English as a second language (ESL) settings than in EFL contexts. Much more importantly, studies like what were reported are mainly addressing attitudes toward L1 use, but they have never approached it in terms of validity perspectives. What adds to the novelty of this study is, first, to approach L1 as a valid device and, second, to address and reinforce its validity in the light of the unitary concept. The theoretical notions of validity shifted from many distinct types to a unified notion with multiple features which closely resemble that of Messick (1989). Hubley and Zumbo (1996) believed that despite theoretical advances, actual validation practice still tends to be based on the classical notion of validity, creating a persistent gap between psychometric theory and research practice. According to Messick (1998), the disjunction between validation practice and validity theory would be solved by technology-based assessment, and both theory and practice would be unified. Messick recommended that all components of validity be incorporated under the concept of construct validity.

This study along with others (e.g., Nakatsuhara & Jaiyote, 2015; Nakatsuhara et al., 2018; Zhou, 2016) employed Messick's framework to lead validation practice. This framework supported our findings in the validation of performance assessment and filled the existing gap between validation practice and validity theory.

Method

Participants

Ninety-seven Iranian intermediate EFL learners (18–25 years old) were selected conveniently from Islamic Azad University (Science and Research Branch) in Tehran province. In a bid to compensate for the sampling issue on one hand and the relative sample size limitation on the other, the mixed-methods research design, in the form of two complementary phases, was followed. So, the integration of the qual-quant trend added to the depth and vigor of the findings. In the quantitative phase, 90 participants were selected based on their performance in Cambridge Preliminary English Test (PET), 2015. The

selected participants' second language (L2) performance was assessed through L1-based and L2-based elicitations techniques. They completed two researcher-made perceptions questionnaires. For the qualitative phase, 15 intermediate English language learners as the focus group of the study were selected through a convenience sampling method ($N=15$: $n=8$ for the L1 group and $n=7$ for the L2 group). According to Denscombe (2007, p. 115), a "focus group consists of a small group of people, usually between six and nine in number." Believing that a focus-group interview provides a setting for the relatively homogeneous group to reflect on the questions asked by the interviewer, the selected participants were interviewed.

Instrumentation

Two researcher-made Likert-scale perception questionnaires, constructed based on the results obtained from focus-group interviews and literature review, were employed to capture the EFL learners' perceptions of L1-based and L2-based elicitation techniques. To this end, after reviewing the existing literature and following the procedures for focus-group interviews offered by Elliot (2005), an attempt was made to execute such an interview for the qualitative phase of the study. Therefore, a focus group was first defined, and the questions were formulated. Then, the focus groups' members comprising 15 individuals ($n=8$ for the L1 group; $n=7$ for the L2 group) were carefully recruited. A homogeneous group of participants comprised the focus group. The following criteria have been considered in the selection of individual groups: age, gender, power, and language proficiency (Elliot, 2005).

Regarding the number of questions posed in the discussion sessions, Elliot considers 12 as a maximum, 10 to be better, and 8 as an ideal number of questions. Reviewing the existing literature on the issues, 8 general questions were developed and raised during the discussion sessions. The participants in the focus groups were not aware of the contents of the questions they were being asked. Some criteria have been set to make sure that the participants have no problems in understanding and responding to the questions posed. The questions were short, to the point, one dimensional, unambiguous, worded precisely, open ended, and non-threatening or embarrassing (Elliot, 2005). The questions are classified into three types: (1) engagement questions, (2) exploration questions, and (3) exit questions. Engagement questions are used to introduce the topic to participants and make them comfortable with the topic of discussion. Exploration questions take the participants to the main part and body of the discussion. Exit questions are used to check if anything was missed in the discussion. Examples are as follows:

Engagement question: Do you ever use your mother tongue when you are doing a task in English (e.g., retelling a story, reading comprehension, listening, speaking, writing a short story)?

Exploration question: What do you do when you cannot understand what your teacher says in English?

Exit question: "Is there anything else you would like to say about your preferences toward L1- and/or L2-based elicitation?"

Considering the abovementioned considerations and criteria, the researcher (moderator) and his assistant conducted the focus groups. The discussion was facilitated by the moderator, and the notes were taken by the assistant. The sessions were also recorded

for further analysis. Finally, the focus-group interviews were conducted in two 90-min sessions (see Appendix for the interview questions and protocols). Following the constant comparison analysis technique, emergent-systematic focus-group design (Onwuegbuzie et al., 2009), the data obtained from focus groups were analyzed. The extracted codes (30 items), categories (5 items), and themes (2 items) were employed to construct the questionnaires. Accordingly, two 30-item Likert-scale questionnaires (see Appendix) were constructed and piloted, the reliability and validity of which were confirmed using Cronbach's alpha and face and content validity by the experts, respectively.

Data collection procedure

Those participants whose scores lay between 140 and 170, based on the Cambridge English Language Assessment rating scale, were identified as qualified individuals to participate in the study. According to the purposes, five kinds of elicitation techniques were employed, namely: (1) asking questions, (2) asking questions combined with pictures, (3) asking questions combined with activities, (4) asking questions combined with texts and dialogues, and (5) asking questions combined with nonverbal language. Defining, synonyms, paraphrasing, forgetting, and asking multiple questions via the participants' L1 (Farsi) were focused. Each technique followed three steps, including opening, questioning, and main activity. In the opening step, the teacher opened the teaching–learning process. In the questioning step, the teacher asked a simple question that was related to the topic of descriptive text, for example, about animals, to elicit the students to talk. In the main activity step, the teacher explained a descriptive text. While the students were retelling a story, which referred to a similar situation and experience of the learners (as a task), the teacher provided them with the definitions of target materials in their L1 (Farsi) and L2 (English), e.g., words, and asked them to come up with the matching word in English. To add a natural taste to the elicitation process, the teacher would pretend to forget the word, the grammatical structure, pronunciation, etc. so that grounds could be intentionally paved for the students to supply the target answer. The teacher would ask questions in Farsi (L1 group) whose answers would require the students to use the target linguistic feature. Some grammar-eliciting techniques such as picture description, conversations, readings, retelling stories, and examples were employed, and the required explanations were also provided by shifting to the learners' L1. Headlines, words, pictures, proverbs, personal notes, free writing, etc. were also provided as a tool for eliciting the learners' ideas. As a formative performance assessment, three similar speaking tests were conducted with a 1-week interval between the tests. Finally, the piloted questionnaires were distributed among the participants and completed. Before the tests, the necessary explanations about the tests and the objectives had been given to the participants.

Data analysis

Researchers working on focus groups do not have a fixed and single framework for analyzing the qualitative data obtained from the focus-group discussion session. But some qualitative data analysis techniques have been identified as appropriate for analyzing these types of data. One of the frameworks suggested by Leech and Onwuegbuzie (2008) encompasses several analytical techniques, including constant comparison analysis, classical content analysis, keywords in context, and discourse

analysis. Following the constant comparison analysis technique, emergent-systematic focus-group design (Onwuegbuzie et al., 2009), the data obtained from the focus groups were analyzed.

Table 1 summarizes the extracted codes, categories, and themes. Initially, in the open coding phase, 30 codes were extracted from the respondents' statements during the interview sessions. These codes were, then, grouped into 5 categories in axial coding stage. Finally, these categories were grouped into 2 themes (L1/L2 use for learning purposes and L1/L2 use for testing purposes) that expressed the content of each of the groups. In doing so, the researchers used two groups to assess if the themes that emerged from one group (L1-based group) also emerged from another group (L2-based group, $n=7$). This assisted the researchers in reaching data saturation.

The analyses of the results of focus-group interviews then were employed to construct L1-based and L2-based perception questionnaires.

Based on the findings obtained from literature and focus-group interviews, the early drafts of the questionnaires were constructed. They were then piloted (L1-based group $n=20$; L2-based group $n=20$) and revised. The language of the questions and the format of the questionnaires were modified. Some existing survey-based studies (e.g., Levine, 2003; Rolin-Ianziti & Varshney, 2008) were reviewed, and some more related items, which met the needs of the current study, were adapted from these studies. For example, one of the items adopted from Rolin-Ianziti and Varshney (2008) was modified, i.e., *I feel more at ease when my teacher uses English* was changed to *I feel more at ease when I am allowed to use my first language in this class*. In revision, the researchers also created four item categories:

✓ *perceptions of L1 Use in General*
✓ *perceptions of L1 use with other students*
✓ *perceptions of L1 use during oral performance*
✓ *perceptions of L1 use during assessment and testing procedures*

One more category was added to the previous ones by the reviewers, i.e., *L1 Use Permission*. Therefore, the final version of the questionnaire was created with five categories. Based on reviewing the mentioned surveys and the results obtained from focus-group interviews, the questions were formulated for each category and added to the ones retained from the pilot version of the questionnaires. The final questions were developed and arranged in a questionnaire on a 5-point Likert scale: 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree). All 30 items were randomized. With the 30-item Likert-scale questionnaires, Cronbach's alpha yielded 0.909 for the L1-based perceptions questionnaire and 0.894 for the L2-based perceptions questionnaire suggesting a satisfactory level of reliability for the instruments.

In addition, the construct validity of the questionnaires was also explored. To do so, five separate exploratory factor analyses were run to investigate the underlying constructs of the five components of the L1-based and L2-based perception questionnaires. Based on the results, it was concluded that the correlation matrices were not singular. There are perfect correlations among all variables of the L1-based and L2-based perception questionnaires.

Table 1 The extracted codes, categories, and themes for the role of L1 and L2 using the constant comparison analysis technique

Themes	Categories	Codes
L1 use for learning purposes	**L1 and L2 Use Permission**	I like it when my teacher allows me to use my first language (Farsi)/(English) in this class
		I believe that both English and Farsi can be used in this class
		Students should be allowed to use their first/second language (Farsi)/(English) in the class
		I believe that to learn English better, students should use English all the time in this class
		I feel more at ease when I am allowed to use my first/second language (Farsi)/(English) in this class
	L1 and L2 Use in General	I think that translating words/sentences/ideas that I do not know helps me while reading texts in English in this class
		Annotating texts in my first/second language helps me understand the texts at a deeper level
		I like translating texts in English into my first language (Farsi) because it helps me learn better
		I prefer writing in my first/second language (Farsi)/(English)
		I think new grammatical points should be presented in my first/second language (Farsi)/(English)
		New vocabulary items should be presented in English and Farsi equivalents
	L1 and L2 Use with Others	I feel that I can relate to my classmates when I use my first/second language (Farsi)/(English) in this class
		I prefer to work in groups with classmates who speak my first/second language (Farsi)/(English) in this class
		I like using my first/second language (Farsi)/(English) with other students to discuss texts assigned in this class that I find challenging
		I prefer to use my first/second language (Farsi)/(English) with my fellow classmates outside of class
		Using my first/second language (Farsi)/(English) helps me communicate better in this class
		I can express my ideas better when I am able to discuss articles with my classmates in my first/second language (Farsi)/(English)

Table 1 (continued)

Themes	Categories	Codes
L1 Use for uesting purposes	**L1 and L2 Use in Oral Performance and Assignments**	I use my first/second language (Farsi)/(English) to help me write essays or assignments for this class
		I believe that translating Farsi into English to retell a story is a good way to learn how to speak in English
		I like to brainstorm/freewrite in my first/second language (Farsi)/(English) while planning my assignments for this class
		I can express my ideas better when I am allowed to write and speak in my first/second language (Farsi)/(English) in this class
		I believe that students should speak in English during group activities in this class
		I like to use my first/second language (Farsi)/(English) to see if I understand the topics/materials discussed in this class
		I like to use an English-to-Farsi dictionary to look up new words when I am reading texts for this class
	L1 and L2 Use in Assessment and Testing	In testing sessions, I prefer to have the questions in my first/second language (Farsi)/(English)
		In testing sessions, while I retell a story in English, it is helpful when the teacher shifts to Farsi for guiding me to the rest of the story
		The teacher's shifting into Farsi does not distract me when I am doing a task in testing sessions
		Depending on the type of tasks (oral, written, fill-in, multiple choice, etc.), only L1 (Farsi) can be used for assessing L2 language
		Depending on the type of tasks (oral, written, fill-in, multiple choice, etc.), only L2 (English) can be used for assessing L2 language
		Depending on the type of tasks (oral, written, fill-in, multiple choice, etc.), both L1 (Farsi) and L2 (English) can be used for assessing L2 language

Results

Table 2 displays Cronbach's alpha reliability for the L1 and L2 questionnaires piloted on a group of 40 students (20 students for L1 and 20 students for L2). The reliability indices for the subsections of the questionnaire for L1 and L2 use were as follows: permission ($L1 = 0.733;\ L2 = 0.874$), general ($L1 = 0.800;\ L2 = 0.611$), others ($L1 = 0.847;\ L2 = 0.693$), oral performance ($L1 = 0.733;\ L2 = 0.874$), and assessment

Table 2 Reliability statistics of L1 and L2 questionnaires (pilot study)

L1 & L2 use	L1	L2	No. of items
Permission	0.733	0.874	5
General	0.800	0.611	6
Others	0.847	0.693	6
Oral performance	0.791	0.611	7
Assessment	0.213	0.280	6
Total	0.909	0.894	30

Table 3 Item-total statistics for L1 and L2 questionnaires

	Scale mean if item deleted		Scale variance if item deleted		Corrected item-total correlation		Cronbach's alpha if item deleted	
	L1	L2	L1	L2	L1	L2	L1	L2
Q1	71.93	112.79	121.775	111.828	0.572	.057	0.899	0.896
Q2	72.48	112.98	124.987	98.702	0.571	0.763	0.900	0.883
Q3	71.86	112.85	120.662	101.106	0.561	0.691	0.900	0.886
Q4	72.50	112.46	124.988	108.126	0.518	0.358	0.901	0.893
Q5	72.24	112.60	118.283	110.585	0.790	0.223	0.895	0.894
Q6	72.40	112.62	123.418	109.516	0.570	0.322	0.900	0.893
Q7	72.62	112.92	119.998	99.184	0.631	0.691	0.898	0.885
Q8	71.86	113.50	127.247	102.468	0.270	0.512	0.906	0.890
Q9	72.36	113.15	121.943	103.446	0.642	0.449	0.898	0.891
Q10	70.40	112.40	136.198	105.819	−0.183	0.589	0.914	0.889
Q11	72.38	112.71	124.388	109.615	0.591	0.292	0.900	0.893
Q12	71.95	112.69	123.266	109.156	0.581	0.270	0.899	0.894
Q13	72.02	113.00	126.268	104.596	0.467	0.490	0.902	0.890
Q14	72.14	113.27	122.564	102.670	0.706	0.622	0.898	0.887
Q15	72.19	112.67	122.597	103.376	0.675	0.582	0.898	0.888
Q16	72.31	112.73	122.512	98.712	0.779	0.650	0.897	0.886
Q17	72.17	112.94	121.459	108.102	0.747	0.342	0.897	0.893
Q18	70.33	112.29	131.545	105.871	0.105	0.532	0.907	0.890
Q19	72.52	112.87	132.060	107.218	0.106	0.455	0.906	0.891
Q20	72.19	112.90	125.865	106.563	0.429	0.465	0.902	0.891
Q21	72.19	112.85	124.060	101.914	0.575	0.663	0.900	0.886
Q22	72.60	112.33	122.539	103.376	0.592	0.630	0.899	0.887
Q23	72.52	112.21	120.792	106.722	0.605	0.546	0.899	0.890
Q24	72.26	112.85	117.759	101.659	0.862	0.608	0.894	0.887
Q25	72.40	112.71	123.954	104.551	0.565	0.653	0.900	0.888
Q26	72.14	112.90	120.369	102.436	0.775	0.660	0.896	0.887
Q27	70.79	113.21	143.587	102.381	−0.590	0.536	0.919	0.889
Q28	72.48	114.81	122.158	117.262	0.627	−0.329	0.899	0.906
Q29	72.14	113.71	127.150	105.785	0.261	0.288	0.906	0.896
Q30	70.81	113.04	135.573	115.147	−0.172	−0.181	0.911	0.905

Table 4 Reliability statistics of L1 and L2 questionnaires and their components

	Cronbach's alpha		No. of items	
	L1	L2	L1	L2
Permission	0.845	0.871	4	4
General	0.754	0.645	6	6
Others	0.803	0.652	6	6
Oral	0.815	0.674	7	7
Testing	0.675	0.667	4	3
Total	0.933	0.916	27	28

(*L1 = 0.213; L2 = 0.280*). The total questionnaires enjoyed reliability indices of 0.909 and 0.894, respectively.

Table 3 displays the item-total correlation indices for the L1 and L2 questionnaires. The results indicated that items 10, 27, and 30 had negative item-total correlations for the L1 questionnaire and items 28 and 30 for the L2 questionnaire. These items were discarded for the final reliability and construct validity estimations.

Table 4 displays the reliability indices for the L1 and L2 questionnaires after removing items that had negative contributions to the total data. The overall questionnaire had a reliability of 0.933 for the L1 and 0.916 for the L2 questionnaire. The reliability for the L1 questionnaire subsections ranged from 0.675 for testing to 0.845 for permission and for the L2 questionnaire subsections it ranged from 0.645 for general to 0.871 for permission.

Five separate exploratory factor analyses were run to investigate the underlying constructs of the five components of the L1 and L2 questionnaires.

Table 5 displays the KMO, Bartlett's test of sphericity, and determinant values for the five subsections of the questionnaires. The sample size was adequate for running factor analyses. The KMO indices were all higher than 0.60 which is the minimum acceptable value (Field, 2009) except for L1 Use in General, L2 use in Oral Performance, and L2 Use in Assessment. Bartlett's tests of sphericity were all significant indicating that the correlation matrices were significantly different from an identity one. The determinant values were all higher than 0.001. Based on these results, it can be concluded that the correlation matrices were not singular, and there are perfect correlations among all variables.

The *L1 Use for Permission* was measured through items 2, 3, 10, and 13. The results of exploratory factor analysis (Table 6) indicated that these four items load under a single factor accounting for 70.18% of the total variance. The *L2 Use for Permission* was measured through items 2, 3, 7, 10, and 13. The results of exploratory factor analysis (Table 6) indicated that these four items load under a single factor accounting for 66.63% of the total variance.

Table 7 displays the factor loadings of the four items measuring L1 Use Permission and five items measuring L2 Use Permission. All factor loadings were higher than 0.50. That is to say, all items had large contributions to their construct.

The L1 Use in General was measured through items 8, 12, 15, 19, 22, and 23. The results of exploratory factor analysis (Table 8) indicated that these six items load under

Table 5 KMO and Bartlett's test for L1 and L2 questionnaires

		Permission		General		Others		Oral		Testing	
		L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
KMO[a]		0.812	0.772	0.585	0.713	0.685	0.639	0.691	0.560	0.643	0.508
BTS[b]	χ^2	71.709	138.746	99.106	41.694	113.569	55.536	134.820	83.150	30.526	52.163
	df	6	10	15	15	15	15	21	21	6	6
	Sig	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
D[c]		0.158	.044	.075	0.389	.051	0.284	.028	0.150	0.456	0.312

Note: [a]Kaiser-Meyer-Olkin measure of sampling adequacy

[b] Bartlett's test of sphericity

[c] Determinant

Table 6 Total variance explained for L1 and L2 Use for Permission

Component	Initial eigenvalues						Extraction sums of squared loadings					
	Total		% of variance		Cumulative %		Total		% of variance		Cumulative %	
	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
1	2.807	3.332	70.181	66.630	70.181	66.630	2.807	3.332	70.181	66.630	70.181	66.630
2	0.522	0.659	13.039	13.189	83.220	79.820						
3	0.405	0.632	10.135	12.630	93.355	92.450						
4	0.266	0.250	6.645	4.997	100.000	97.447						
5	–	0.128	–	2.553	–	100.000						

Table 7 Component matrix for L1 and L2 Use for Permission

	Component	
	1	1
	L1	L2
Q3	0.893	0.852
Q13	0.864	0.683
Q7	0.809	0.897
Q2	0.780	0.914
Q10	–	0.707

two factors accounting for 68.72% of the total variance. The L2 Use in General was measured through items 8, 12, 15, 19, 22, and 23. The results of exploratory factor analysis (Table 9) indicated that these six items load under two factors accounting for 56.08% of the total variance.

Table 10 displays the factor loadings of the six items measuring L1 and L2 Use in General. Item 19 had its loading under the second factor. Item 15 also partially loaded under the second factor. All factor loadings were higher than 0.50. That is to say, all items had large contributions to their construct. Regarding the L2, all items loaded under the first factor, while items 22, 23, 12, and 15 had loadings on both factors.

Table 8 Total variance explained for L1 and L2 Use in General

Components of L1 and L2	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.002	50.030	50.030	3.002	50.030	50.030	2.986	49.763	49.763
2	1.121	18.691	68.721	1.121	18.691	68.721	1.138	18.958	68.721
3	0.845	14.087	82.809						
4	0.576	9.593	92.402						
5	0.308	5.138	97.540						
6	0.148	2.460	100.000						
1	2.314	38.560	38.560	2.314	38.560	38.560	1.786	29.775	29.775
2	1.051	17.524	56.084	1.051	17.524	56.084	1.579	26.309	56.084
3	0.973	16.215	72.298						
4	0.664	11.063	83.362						
5	0.538	8.962	92.324						
6	0.461	7.676	100.000						

Table 9 Component matrix for L1 and L2 Use in General

	Components of L1		Components of L2	
	1	2	1	2
Q23	0.874	–	0.333	0.722
Q22	0.834	–	0.613	0.480
Q15	0.792	0.367	0.468	0.576
Q12	0.712	–	−0.312	0.700
Q8	0.616	–	0.793	–
Q19	–	0.968	0.595	–

Table 10 Total variance explained for L1 and L2 Use with Others

Components of L1 and L2	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.152	52.541	52.541	3.152	52.541	52.541	2.562	42.708	42.708
2	1.371	22.857	75.398	1.371	22.857	75.398	1.961	32.690	75.398
3	0.616	10.271	85.669						
4	0.421	7.016	92.685						
5	0.271	4.524	97.209						
6	0.167	2.791	100.000						
1	2.276	37.933	37.933	2.276	37.933	37.933	2.026	33.774	33.774
2	1.327	22.121	60.054	1.327	22.121	60.054	1.577	26.280	60.054
3	0.817	13.618	73.673						
4	0.748	12.466	86.139						
5	0.553	9.220	95.360						
6	0.278	4.640	100.000						

Table 11 Component matrix for L1 and L2 Use with Others

	Components of L1		Components of L2	
	1	2	1	2
Q4	0.913	–	–	0.717
Q6	0.881	–	–	0.526
Q17	0.772	0.457	0.621	0.310
Q16	0.563	0.550	0.882	–
Q21	–	0.924	0.886	–
Q1	–	0.767	–	0.805

The L1 use with others was measured through items 1, 4, 6, 16, 17, and 21. The results of exploratory factor analysis (Table 10) indicated that these six items load under two factors accounting for 75.39% of the total variance. The L2 Use with Others was measured through items 1, 4, 6, 16, 17, and 21. The results of exploratory factor analysis (Table 10) indicated that these six items load under two factors accounting for 60.05% of the total variance.

Table 11 displays the factor loadings of the six items measuring L1 Use with Others. Items 4, 6, 17, and 16 loaded under the first factor, while items 1 and 21 had their loadings under the second factor. Items 16 and 17 loaded under both factors. All factor loadings were higher than 0.50 except for item 17's loading on the second factor.

The L1 Use in Oral Performance was measured through items 5, 9, 11, 14, 18, 20, and 24. The results of exploratory factor analysis (Table 12) indicated that these seven items load under two factors accounting for 68.68% of the total variance. The L2 Use in Oral Performance was measured through items 5, 9, 11, 14, 18, 20, and 24. The results of exploratory factor analysis (Table 12) indicated that these seven items load under three factors accounting for 70.69% of the total variance.

Table 13 displays the factor loadings of the seven items measuring L1 Use in Oral Performance. Items 14, 5, 24, 20, and 9 loaded under the first factor, while items 18

Table 12 Total variance explained for L1 and L2 Use in Oral Performance

Components of L1 and L2	Initial eigenvalues			Extraction Sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.525	50.355	50.355	3.525	50.355	50.355	3.219	45.991	45.991
2	1.283	18.326	68.681	1.283	18.326	68.681	1.588	22.690	68.681
3	0.889	12.694	81.375						
4	0.554	7.919	89.293						
5	0.342	4.890	94.183						
6	0.269	3.843	98.026						
7	0.138	1.974	100.000						
1	2.520	36.005	36.005	2.520	36.005	36.005	2.263	32.324	32.324
2	1.350	19.287	55.291	1.350	19.287	55.291	1.505	21.506	53.830
3	1.078	15.404	70.695	1.078	15.404	70.695	1.181	16.866	70.695
4	0.822	11.746	82.442						
5	0.621	8.877	91.319						
6	0.415	5.925	97.243						
7	0.193	2.757	100.000						

and 11 had their loadings under the second factor. Items 9, 24, and 11 loaded under both factors. All factor loadings were higher than 0.50 except for secondary item loadings. Table 13 also displays the factor loadings of the seven items measuring L2 Use in Oral Performance. Items 14, 24, 18, and 20 loaded under the first factor, while items 11 and 9 had their loadings under the second factor and item 5 loaded under the last factor. Items 24 and 20 loaded under two factors. All primary factor loadings were higher than 0.50 except for secondary item loadings.

The L1 Use in Assessment was measured through items 25, 26, 28, and 29. The results of exploratory factor analysis (Table 14) indicated that these four items load under a single factor accounting for 51.82% of the total variance. The L2 Use in Assessment was measured through items 25, 26, 28, and 29. The results of exploratory factor analysis (Table 14) indicated that these four items load under a single factor accounting for 53.77% of the total variance.

Table 15 displays the factor loadings of the four items measuring L1 and L2 Use in Assessment. The table shows that all items loaded under a single factor, and the obtained values were higher than 0.50.

The students' perceptions of L1-based elicitation were explored using a researcher-made questionnaire, namely, the L1-based perceptions questionnaire. Students' perceptions were investigated in five categories of L1 Use Permission, L1 Use in General, L1 Use with Others, L1 Use in Oral Performance and Assignments, and L1 Use in Assessment and Testing (Table 16).

Regarding the students' *L1 Use Permission* in speaking, 27.142% (12 out of 42 individuals) of the respondents believe that L1 should be used in speaking classes, while 57.14% of the respondents (24 out of 42 individuals) believe that L1 use should not be allowed in speaking classes. The rest of the respondents (15.712%, 6 out of 42 individuals) were neutral.

Table 13 Rotated Component Matrix for L1 and L2 Use in Oral Performance

	Components of L1		Components of L2		
	1	2	1	2	3
Q14	.885	--	.931	--	--
Q5	.848	--	--	--	.918
Q24	.803	.393	.714	.393	--
Q20	.729	--	.620	--	.544
Q9	.577	.484	--	.796	--
Q18	--	.776	.693	--	--
Q11	.426	.699	--	.843	--

Table 14 Total Variance Explained for L1 and L2 Use in Assessment

Component	Initial Eigenvalues						Extraction Sums of Squared Loadings					
	Total		% of Variance		Cumulative %		Total		% of Variance		Cumulative %	
	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
1	2.073	2.151	51.827	53.775	51.827	53.775	2.073	2.151	51.827	53.775	51.827	53.775
2	.964	.927	24.098	23.176	75.925	76.951						
3	.543	.697	13.586	17.432	89.511	94.383						
4	.420	.225	10.489	5.617	100.000	100.000						

Table 15 Component Matrix for L1 and L2 Use in Assessment

	Component of L1	Component of L2
	1	1
Q26	.826	.926
Q28	.754	.631
Q25	.656	.696
Q29	.626	.642

Table 16 Percent Mean for the Five Categories of the L1 Perceptions Questionnaire

Components	Strongly Agree *n* (Percent Mean)	Agree *n* (Percent Mean)	Neutral *n* (Percent Mean)	Disagree *n* (Percent Mean)	Strongly Disagree *n* (Percent Mean)
L1 Use Permission	3 (6.19)	9 (20.952)	6 (15.712)	19 (45.714)	5 (11.426)
L1 Use in General	0 (0)	4 (9.523)	8 (18.65)	25 (59.126)	5 (12.696)
L1 Use with Others	1 (0.793)	3 (5.95)	6 (15.475)	29 (69.445)	3 (6.348)
L1 Use in Oral Performance and Assignments	2 (4.421)	7 (16.664)	6 (13.264)	25 (59.525)	2 (6.12)
L1 Use in Assessment and Testing	1 (1.586)	13 (31.743)	5 (12.301)	20 (46.825)	3 (7.538)

$N = 42$

Table 17 Percent Mean for the Five Categories of the L2 Perceptions Questionnaire

Components	Strongly Agree *n* (Percent Mean)	Agree *n* (Percent Mean)	Neutral *n* (Percent Mean)	Disagree *n* (Percent Mean)	Strongly Disagree *n* (Percent Mean)
L2 Use Permission	12 (24.166)	26 (54.168)	7 (17.416)	3 (6.248)	0 (0)
L2 Use in General	14 (29.858)	25 (53.123)	6 (12.151)	3 (4.86)	0 (0)
L2 Use with Others	10 (21.52)	31 (65.97)	4 (8.33)	2 (3.82)	1 (0.34)
L2 Use in Oral Performance and Assignments	9 (19.34)	31 (63.69)	5 (11.30)	3 (5.65)	0 (0)
L2 Use in Assessment and Testing	6 (11.80)	21 (44.44)	9 (19.09)	10 (20.83)	2 (3.82)
$N = 48$					

Considering the students' *L1 Use in General*, 9.523% (4 out of 42 individuals) of the respondents believe that L1 should be used in general speaking classes, especially in translating words or sentences during reading, writing, or speaking activities, while 71.822% of the respondents (30 out of 42 individuals) believe that L1 should not be used in general speaking classes. The other respondents (18.65%, 8 out of 42 individuals) were neutral. In terms of the students' *L1 Use with Others*, 6.743% (4 out of 42 individuals) of the respondents prefer to use L1 when they want to communicate with each other inside and outside of the classroom, while 75.793% of the respondents (32 out of 42 individuals) do not prefer to use L1 when they want to communicate with each other inside and outside of the classroom. The rest of the respondents (15.475%, 6 out of 42 individuals) are neutral. Regarding the students' *L1 Use in Oral Performance and Assignments*, 21.085% (9 out of 42 individuals) of the respondents prefer to use L1 in doing their assignments or performing orally in class, while 65.645% of the respondents (27 out of 42 individuals) do not prefer to use L1 in doing their assignments or performing orally in the class. The other respondents (13.264%, 6 out of 42 individuals) are neutral. Considering the students' *L1 Use in Assessment and Testing*, 33.329% (14 out of 42 individuals) of the respondents prefer to use L1 in assessment and testing sessions, while 54.363% of the respondents (23 out of 42 individuals) do not prefer to use L1 in assessment and testing sessions. The rest of the respondents (12.301%, 5 out of 42 individuals) were neutral.

Table 17 summarizes the findings of the L2 perception questionnaire. In terms of the students' *L2 Use Permission* in speaking, 78.334% (38 out of 48 individuals) of the respondents believe that L2 should be used in speaking classes.

Regarding the students' *L2 Use in General*, 82.981% (39 out of 48 individuals) of the respondents believe that L2 should be used in general speaking classes, especially translating words or sentences during reading, writing, or speaking activities. Considering the students' *L2 Use with Others*, 87.49% (41 out of 48 individuals) of the respondents prefer to use L2 when they want to communicate with each other inside and outside of the classroom. In terms of the students' *L2 Use in Oral Performance and Assignments*, 83.03% (40 out of 48 individuals) of the respondents prefer to use L2 in doing their assignments or performing orally in class. Regarding the students' *L2 Use in Assessment and Testing*, 56.24% (27 out of 48 individuals) of the respondents prefer to use L2 in assessment and testing sessions.

Discussion

Despite the existing inherent difficulties in construct validity in defining abilities in terms of the abstract nature of constructs (Kane, 2013), Messick believes that construct validity is the central component in validation work and encompasses the five sources of evidence (as discussed before) relevant to the validation of the interpretation and use of the score of an instrument. Five separate exploratory factor analyses were run to investigate the underlying constructs of the five components of the L1-based and L2-based perceptions questionnaires. The reliability of all five components was examined using Cronbach's alpha coefficients. The *Item-Total Statistics* were also examined for corrected item-total correlation. Messick's unitary concept of validity confirms the notion of perceptual validity in the present study since the interpretation of the stakeholders like the learners by itself is a source of inference and then affects the validity as a whole. Other scholars in language testing (e.g., Bachman & Palmer, 2010; Gokturk Saglam & Tsagari, 2022; Nguyen & Habok, 2021; Shohamy, 2001) have also supported collecting stakeholders' perspectives for validation. Moreover, Im et al. (2019) emphasized that actual language users in the target domain can be involved in evaluating test items during the validation.

Hence, after the examination and verification of the construct validity of different categories of L1-based and L2-based questionnaires and running the statistical analyses, the findings of the present study show that 27.142% of the respondents of the L1-based questionnaire believe that L1 should be used in speaking classes, while 57.14% of the respondents of the questionnaire believe that L1 use should not be allowed in speaking classes. In other words, the majority of respondents prefer to use their L2 in speaking classes. On the other hand, the majority of the respondents to the L2-based questionnaire (78.334%) believe that L2 should be used in speaking classes rather than L1. Therefore, both groups of respondents favor L2 use in their speaking classes.

In response to the second category of the questionnaire, 9.523% of the respondents of the L1-based questionnaire believe that L1 should be used in general speaking classes, especially for translating words or sentences during reading, writing, or speaking activities, while 71.822% of the respondents believe that L1 should not be used in general speaking classes. Therefore, the majority of the respondents believe that L2 should be the prior language in general speaking classes, especially for translating words or sentences during reading, writing, or speaking activities. On the other hand, the majority of the respondents to the L2-based questionnaire (82.981%) also believe that L2 should be used in general speaking classes for the same activities. Therefore, in the second category of the questionnaire, both L1 and L2 groups prefer to use their L2.

In response to the category of *L1 Use with Others*, 6.743% of the respondents of the L1-based questionnaire prefer to use L1 when they want to communicate with each other inside and outside of the classroom, while 75.793% of the respondents do not prefer to use L1 in the same situations. It means that more than two-thirds of the respondents prefer to use L2 when they want to communicate with each other inside and outside of the classroom. On the other hand, the majority of the respondents to the L2-based questionnaire (87.49%) prefer to use L2 when they want to communicate with each other inside and outside of the classroom, meaning that both groups favor L2 for communication inside and outside of the classroom.

In investigating the *L1 Use in Oral Performance and Assignments*, 21.085% of the respondents of the L1-based questionnaire prefer to use L1 in doing their assignments or performing orally in class, while 65.645% of the respondents do not prefer to use L1 in doing their assignments or performing orally in the class, meaning that the majority of the respondents prefer to use L2 in doing their assignments or performing orally in class. On the other hand, the majority of the respondents to the L2-based questionnaire (83.03%) prefer to use L2 in doing their assignments or performing orally in class.

In exploring the learners' perceptions of *L1 Use in Assessment and Testing*, 33.329% of the respondents of the L1-based questionnaire prefer to use L1 in assessment and testing sessions, while 54.363% of the respondents do not prefer to use L1 in assessment and testing sessions. It shows that more than half of the respondents prefer to use L2 in assessment and testing sessions. On the other hand, the majority of the respondents to the L2-based questionnaire (56.24%) prefer to use L2 in assessment and testing sessions.

As a final point, the results show that in response to the five categories of both L1-based and L2-based questionnaires, i.e., *L1 and L2 Use Permission*, *L1 and L2 Use in General*, *L1 and L2 Use with Others*, *L1 and L2 Use in Oral Performance and Assignments*, and *L1 and L2 Use in Assessment and Testing*, the majority of the respondents prefer to use their L2 in their speaking classes rather than their L1. The use of L1 is well documented in studies in the literature (e.g., Afzal, 2012; Cook, 2008; Littlewood & Shufang, 2022; Nazary, 2008; Saito & Ebsworth, 2004; Shin et al., 2019); it is therefore somewhat surprising that majority of the students favor L2 in their English classes for various functions. We suspect that the overemphasis on the use of the L2 by other subject teachers may encourage the students to think that the use of L1 might not be as much useful as their L2 since the students in the L1-based elicitation group outperformed their counterparts in the L2-based elicitation group in doing their language performances (Mohammadi Darabad et al., 2021). Moreover, Shin et al. (2019) reviewed recent articles on L1 use in EFL classrooms. Contrary to the findings of the present study, wherein the preference for L2 was emphasized more, the valuable role of L1 in L2 learning was acknowledged by the majority of the students and teachers in these studies (cf. Chiou, 2014; Liu & Zeng, 2015; Mohebbi & Alavi, 2014; Tsagari & Diakou, 2015). The data obtained in Mohebbi and Alavi's (2014) study showed that L1 was mainly used by the L2 teachers to provide feedback, teach new vocabulary, explain grammar, build rapport, manage the class, give individual help to learners, and save time in lengthy tasks explanations. In Macaro and Lee's (2013) study, the participants' attitudes toward L1 use were explored. The results of their study indicated that both children and college students considered L1 use a valuable tool in their EFL classes. On the other hand, when L2-only instruction was taken into consideration, adult learners were more contented than children. In the present study, therefore, along with Macaro and Lee's, age could be a prevailing factor for the students' perceptions of L1 or L2.

In a study conducted by Ferrando (2013), in examining the students' perceptions of L1 use in a multilingual setting, the students neither agreed nor disagreed with using their first language in classrooms. Many of the questionnaire items regarding the *L1 Use with Classmates* were marked "neutral" by the students, while in the present study, the students preferred L2 use with others. In terms of *Language Use in General*, in contrast with the present study, the majority of the students marked *Strongly Agreed*

choice representing a strong pattern of agreement. In terms of whether L1 use should be allowed in the classroom or not, like in the present study, about half of the students believed that L2 use should be allowed in the classroom.

In exploring the students' perceptions of L1 and/or L2 use in EFL classrooms, the students' perceptions of L1 and L2 use can be attributed to the following reasons: (1) English language (L2) is considered the medium of instruction in the classroom since, in the students' views, it provides the maximum exposure to the target language, and the students have a chance to connect with other students to encourage them to use the target language in the class; (2) while the results show a strong desire to use L2 in the classroom, there is a consensus among some students that they should be allowed to use their first language at least in testing and assessment sessions. They believe that using L1 may help them with the target language tasks. The findings in the present study show that students prefer to use L2 (English) rather than their L1 (Farsi). This finding is similar to Ferrando's study and some previous studies (e.g., Al Sharaeai, 2012; Rivers, 2011; Turnbull, 1999) that examined L1 use in EFL contexts where all students speak the same L1. Along with other studies, including Ferrando's (2013), Nazary's (2008) study also acknowledged that university students in Iran are reluctant to use their L1 in English language situations and reject it strongly for the sake of better exposure to L2. While Cook (2008) asserted that L2 teachers might fall back on learners' L1 for conveying meaning, managing the class, giving instructions for teaching activities, and testing, the obtained data, which were collected from three different language proficiency levels (including beginning, intermediate, and advanced) indicated that the importance and effectiveness of L1 use were not supported by the majority of students. However, as Littlewood and Shufang (2022) elaborated, it is better to adopt a more balanced view. There is a consensus among teachers that in contexts where the teacher is the learners' main or only source of L2 input, it is important to create a classroom in which the L2 is dominant. On the other hand, they see that a policy of total exclusion does not work in practice and indeed question whether it is advisable. After all, students' first language is a rich source of support in understanding the new language and a powerful instrument for creating learning situations.

Conclusion and implications

The examination and exploration of the L1-based and L2-based questionnaires endorsed their construct validity. The results from this validated questionnaire revealed that the EFL learners in the Iranian context believed in the effectiveness of their L2 (English) use in their EFL classrooms compared with their L1 use, and more than half of them were reluctant to use their L1 (Farsi), especially for learning purposes (L1 Use Permission, L1 Use in General, and L1 Use with Others) and testing purposes (L1 Use in Oral Performance and Assignments and L1 Use in Assessment and Testing). However, some studies showed that careful use of L1 is beneficial and facilitative in L2 learning, at the same time stressing that too much use of the L1 might reduce the students' contact with the L2 and deprive them of practicing it. The study found that the students' use of L2 is superior to their use of L1 in the five mentioned categories. However, there were some

learning instances, including giving instructions, introducing new words, and explaining grammatical rules where the use of L1 was helpful.

According to the obtained results and the learners' perceptions of L1-based and L2-based elicitation techniques, the use of the learners' L2, in their views, can be more facilitative than their L1 for understanding their L2 both for learning and testing aims. Thus, the teachers should devote more time to using L2 in their classes and, at the same time, take advantage of L1 use when needed. Similar studies at different language proficiency levels are suggested to be conducted by other researchers to explore more viewpoints regarding various aspects of L1 use in English learning classes.

Abbreviations
KMO Kaiser Meyer Olkin

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40468-023-00218-4.

Additional file 1. L1-Based PerceptionQuestionnaire. L2-Based Perception Questionnaire

Acknowledgements
We are grateful to our colleagues and students for their participation and their invaluable feedback on the manuscript.

Authors' contributions
We are the sole contributors of all the content in this manuscript. The authors read and approved the final manuscript.

Funding
The authors received no funding for this research.

Availability of data and materials
The data are available upon request from the authors.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 18 November 2022 Accepted: 19 January 2023
Published online: 22 February 2023

References

Afzal, S. (2012). Relationship between providing Persian equivalents of English adjectives and Iranian EFL learners' active vocabulary. *International Journal of English Linguistics,2*(1), 231–238. https://doi.org/10.5539/ijel.v2n1p231
Al Sharaeai, W. A. A. (2012). Students' perspectives on the use of L1 in English classrooms. *Graduate Thesis and Dissertations*. 12898. Retrieved from: https://doi.org/10.31274/etd-180810-800
Alshammari, M. M. (2011). The use of the mother tongue in Saudi EFL classrooms. *Journal of International Education Research*, *7*(4), 95–102. https://doi.org/10.19030/jier.v7i4.6055
Auerbach, E. (1993). Re-examining English only in the ESL classroom. *TESOL Quarterly,27*(1), 9–32. https://doi.org/10.2307/3586949
Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
Chiou, B. (2014). Rethinking the role of L1 in the EFL classroom. *English Teaching & Learning, 38*(4), 53–78. 10.6330ETL.2014.
Cook, V. (2008). Second language learning and language teaching. UK: Hodder Education.
Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd ed.). New York: McGraw-Hill.
Eckman, F. (2014). Second language phonology. In S. M. Gass, & A. Mackey (eds). *The Routledge Handbook of Second Language Acquisition* (pp. 91–105). New York: Routledge.
Elliot, A. (2005). *Guidelines for conducting a focus group.* Retrieved, Aprill 04, 2017, from: https://assessment.aas.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf
Ellis, R. (1984). C*lassroom Second Language Development*. Oxford: Pergamon
Ferrando, J. L. (2013). *Generation 1.5 students' perceptions of L1 use in college-level multilingual composition classroom*. A thesis presented to the faculty of the Department of English California State University, Sacramento.
Field, A. (2009). *Discovering statistics using SPSS: Introducing statistical method* (3rd ed.). Thousand Oaks: Sage.
Gaillard, S., & Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: The elicited imitation task. *Language Learning,66*(2), 419–447. https://doi.org/10.1111/lang.12157

Gass, S. (2018). SLA elicitation tasks. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (eds.), The Palgrave Handbook of Applied Linguistics Research Methodology (pp. 313–337). English Language Center, Michigan State University, East Lansing, MI, USA. https://doi.org/10.1057/978-1-137-59900-1_15

Gokturk Saglam, A. L., & Tsagari, D. (2022). Evaluating perceptions towards the consequential validity of integrated language proficiency assessment. *Languages,7*(65), 2–18. https://doi.org/10.3390/languages7010065

Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology, 123*(3), 203–215. https://doi.org/10.1080/00221309.1996.9921273

Im, G. H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: Their limitations and future directions for validation research. *Language Testing in Asia,9*(14), 1–26. https://doi.org/10.1186/s40468-019-0089-4

Jee-Young Shin, L., Quentin, D., & Yunkyeong, Ch. (2019). An updated review on use of L1 in foreign language classrooms. *Journal of Multilingual and Multicultural Development,24*, 1–14. https://doi.org/10.1080/01434632.2019.1684928

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement,50*(1), 1–73.

Khati, A. (2011). When and why of mother tongue use in English classrooms. *Journal of NELTA,16*(1–2), 42–51. https://doi.org/10.3126/nelta.v16i1-2.6128

Lardiere, D. (2014). Linguistic approaches to second language morphosyntax. In S. M. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 106–126). New York: Routledge.

Leech, N. L., & Onwuegbuzie, A. J. (2008). Qualitative data analysis: A compendium of techniques for school psychology research and beyond. *School Psychology Quarterly,23*, 587–604. https://doi.org/10.1037/1045-3830.23.4.587

Levine, G. (2003). Target language use, first language use, and anxiety: Report of a questionnaire study. *Modern Language Journal,87*, 343–364. https://doi.org/10.1111/1540-4781.00194

Littlewood, W., & Shufang, W. (2022). The role of learners' first language in the foreign language classroom. *Language Teaching Research Quarterly, 31*, 174–182. https://doi.org/10.32038/ltrq.2022.31.12

Liu, Y., & Zeng, A. P. (2015). Loss and gain: Revisiting the roles of the first language in novice adult second language learning classrooms. *Theory and Practice in Language Studies 5*(12), 2433–2440. https://doi.org/10.17507/tpls.0512.01

Macaro, E., & Lee, J. H. (2013). Teacher language background, codeswitching, and English-only instruction: Does age make a difference to learners' attitudes? *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect,47*(4), 717–742. https://doi.org/10.1002/tesq.74

Mahmoudi, L., & Amirkhiz, S. (2011). The use of Persian in the EFL classroom–The case of English teaching and learning at pre-university level in Iran. *English Language Teaching,4*(1), 135–140. https://doi.org/10.5539/elt.v4n1p135

McManus, K. (2022). *Crosslinguistic influence and second language learning*. New York and London: Routledge, Taylor & Francis. https://doi.org/10.4324/9780429341663

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 45*, 35–44. https://doi.org/10.1023/A:1006964925094

Mohammadi Darabad, A., Abbasian, Gh. R., Mowlaie, B., & Rostami Abusaeedi, A. A. (2021). L1-based elicitation as a valid measure of L2 classroom performance assessment. *Issues in Language Teaching, 10*(1), 1–36. https://doi.org/10.22054/ilt.2021.52989.507

Mohebbi, H., & Alavi, S. M. (2014). An investigation into teachers' first language use in a second language learning classroom context: A questionnaire-based study. *Bellaterra Journal of Teaching and Learning Language and Literature,7*(4), 57–73. https://doi.org/10.5565/rev/jtl3.539

Molway, L., Arcos, M., & Macaro, E. (2022). Language teachers' reported first and second language use: A comparative contextualized study of England and Spain. *Language Teaching Research,26*(4), 642–670. https://doi.org/10.1177/1362168820913978

Nakatsuhara, F., & Jaiyote, S. (2015). *Exploring the impact of test-takers' L1 backgrounds on paired speaking test performance: How do they perform in shared and non-shared L1 pairs? Paper presented at the BAAL/CUP Applied Linguistics Seminar*. UK: York St John University. 24-26/06/2015.

Nakatsuhara, F., Taylor, L., & Jaiyote, S. (2018). The role of the L1 in testing L2 English. In C. J. Hall & R. Wicaksono (Eds.), *Ontologies of English Conceptualizing the Language for Learning, Teaching, and Assessment* (pp. 1–18). Cambridge: Cambridge University Press.

Nazary, M. (2008). The role of L1 in L2 acquisition: Attitudes of Iranian university students. *Novitas Royal,2*(2), 138–153.

Nguyen, S. V., & Habok, A. (2021). Designing and validating the learner autonomy perceptions questionnaire. Heliyon, 7, e06831. https://doi.org/10.1016/j.heliyon.2021. e06831

Onwuegbuzie, A. J., Slate, J. R., Leech, N. L., & Collins, K. M. T. (2009). Mixed data analysis: Advanced integration techniques. *International Journal of Multiple Research Approaches,3*, 13–33. https://doi.org/10.5172/mra.455.3.1.13

Perkins, K., & Jiang, X. (2019). Neuroimaging and reading comprehension. *Journal of Interdisciplinary Studies in Education 8*(2), 74–94. https://doi.org/10.32674/jise.vi0.649

Perkins, K., & Zhang, L. J. (2022). The effect of first language transfer on second language acquisition and learning: From contrastive analysis to contemporary neuroimaging. *RELC Journal,* 1–17. https://doi.org/10.1177/00336882221081894

Plonsky, L., Marsden, E., Crowther, D., Gass, S., & Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research,35*(1), 1–39.

Polio, C. (2014). The acquisition of second language writing. In S. M. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 319–334). New York: Routledge

Ringbom, H., & Jarvis, S. (2009). The importance of cross-linguistic similarity in foreign language learning. In M. H. Long and C. J. Doughty (eds.). *The Handbook of Language Teaching* (pp.106–118). Chichester: Wiley-Blackwell. https://doi.org/10.1002/9781444315783.ch7

Rivers, D. J. (2011). Politics without pedagogy: Questioning linguistic exclusion. *ELT Journal,65*(2), 103–113. https://doi.org/10.1093/elt/ccq044

Rolin-Ianziti, J., & Varshney, R. (2008). Students' views regarding the use of the first language: An exploratory study in a tertiary context maximizing target language use. *The Canadian Modern Language Review,65*(2), 249–273. https://doi.org/10.3138/cmlr.65.2.249

Saito, H., & Ebsworth, M. E. (2004). Seeing English language teaching and learning through the eyes of Japanese EFL and ESL students. *Foreign Language Annals,37*, 111–124. https://doi.org/10.1111/j.19449720.2004.tb02178.x
Schweers, C. W. (1999). Using L1 in the L2 classroom. *English Teaching Forum,37*(2), 6–13.
Sharma, K. (2006). Mother tongue use in English classroom. *Journal of NELTA,11*(1–2), 80–87. https://doi.org/10.3126/nelta.v11i1.3132
Shin, J. Y., Quentin Dixon, L., & Choi, Y. (2019). An updated review on use of L1 in foreign language classrooms. *Journal of Multilingual and Multicultural Development, 41*(5):1–14. https://doi.org/10.1080/01434632.2019.1684928
Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing,18*(4), 373–391. https://doi.org/10.1191/026553201682430094
Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing,11*(2), 99–124. https://doi.org/10.1177/026553229401100202
Shohamy, E., & Stansfield, C. (1991). The Hebrew oral test: An example of international cooperation. *AILA Bulletin,7*, 79–90. https://doi.org/10.5325/j.ctv14gp4vq.10
Shohamy, E., Gordon, C., Kenyon, D. M., & Stanfield, C. W. (1989). The development and validation of a semi-direct test for assessing oral proficiency in Hebrew. *Bulletin of Hebrew Higher Education,4*, 4–9.
Stansfield, C. W., & Kenyon, D. M. (1992). The development and validation of a simulated oral proficiency interview. *Modern Language Journal,76*, 129–141. https://doi.org/10.1111/j.1540-4781.1992.tb01093.x
Tsagari, D., & Diakou, C. (2015). Students' and teachers' attitudes towards the use of the first language in the EFL state school classrooms. *Research Papers in Language Teaching & Learning 6*(1), 86–108. Available online at http://rpltl.eap.gr
Turnbull, M. (1999). Multidimensional project-based teaching in French second language (FSL): A process-product study. *Modern Language Journal,83*, 548–568. https://doi.org/10.1111/0026-7902.00040
Wigglesworth, G., & O'Loughlin K. (1993). An investigation into the comparability of direct and semi-direct versions of an oral interaction test. *Paper presented at the 15th Language Testing Research Colloquium*, Cambridge, August.
Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals,465*(4), 680–704. https://doi.org/10.1111/flan.12063
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing,33*(4), 497–528. https://doi.org/10.1177/0265532215594643
Yao, M. (2011). On attitudes to teachers' code-switching in EFL classes. *World Journal of English Language,1*(1), 19–28. https://doi.org/10.5430/wjel.v1n1p19
Zhou, W. (2016). Investigating the construct validity of communicative proficiency in TEP (oral) at level B. *Journal of Language Teaching and Research, 7*(4), 690–699. https://doi.org/10.17507/jltr.0704.08

Publisher's Note