RAEL: Revista Electrónica de Lingüística Aplicada
Vol./Núm.: 15/1. Año 2016
Enero-diciembre 2016
Páginas: 129-149
Artículo recibido: 29/07/2016
Artículo aceptado: 06/10/2016
Artículo publicado: 31/01/2017
Url: http://www.aesla.org.es/ojs/index.php/RAEL/article/view/287

Towards the Validation of a Scale for Measuring the Load of Form Focus and Meaning Focus of Textbook Activities in Foreign Language Teaching

Hacia la validación de una escala para medir el peso de la atención a la forma y al significado de las actividades de los libros de texto en la Enseñanza de Segundas Lenguas

RAQUEL CRIADO
UNIVERSITY OF MURCIA, SPAIN

Foreign Language Teaching is targeted at the attainment of implicit language knowledge by means of form-focused and meaning-focused instruction (FFI and MFI). This study is aimed at devising and validating a scale with psychometric techniques that measures the degree of FFI and MFI of textbook activities. Specifically, the objectives of this article are twofold. The first objective is to study the concurrent criterion validity of the application of the scale by 5 applied linguists and 5 English as a Foreign Language (EFL) secondary-school teachers on 15 randomly-selected authentic EFL textbooks. The external criterion is the holistic assessment of the FFI and MFI loads of these activities on a continuum from 0 to 10. The second objective is to study the inter-rater reliability scores of both types of assessments (holistic and scale-based). Results show very satisfactory concurrent criterion validity and inter-rater reliability. Future lines of research are suggested.

Keywords: *form-focused instruction; meaning-focused instruction; textbook activities; scale; concurrent criterion validity*

El objetivo último de la Enseñanza de Segundas Lenguas consiste en alcanzar el conocimiento implícito de la lengua a través de la enseñanza centrada en la forma (ECF) y en el significado (ECS). El presente estudio está destinado a diseñar y validar una escala mediante técnicas psicométricas que mida el grado de ECF y ECS de las actividades de libros de texto. Este artículo tiene dos objetivos específicos. El primero persigue estudiar la validez de criterio concurrente derivada de la aplicación de la escala por parte de 5 lingüistas aplicados y 5 profesores de Educación Secundaria de Inglés como Lengua Extranjera (ILE) a 15 actividades auténticas de libros de texto de ILE aleatoriamente seleccionadas. El criterio externo es la evaluación holística de los pesos de ECF y ECS de dichas actividades en un continuo de 0 a 10. El segundo objetivo pretende estudiar la fiabilidad interjueces de las puntuaciones de ambos tipos de evaluaciones (holísticas y basadas en la escala). Los resultados revelan que existe una validez de criterio concurrente y una fiabilidad inter-jueces muy satisfactorias. Finalmente, se proponen futuras líneas de investigación.

Palabras clave: *enseñanza centrada en la forma; enseñanza centrada en el significado; actividades de libros de texto; escala; validez de criterio concurrente*

1. INTRODUCTION

The essential goal of Foreign Language Teaching (FLT) as framed within a communicative and cognitive approach should be the attainment of implicit language knowledge (Ellis, 2013, 2015; DeKeyser, 2015; among others). Implicit language knowledge refers to knowledge of a language that may be accessed instantaneously during spontaneous comprehension or production. It differs from explicit language knowledge, which is conscious knowledge about language forms – grammatical rules, lexicon, pronunciation patterns, pragmatic and sociolinguistic conventions (Akakura, 2012). If explicit knowledge exerts any influence on the optimal attainment of implicit knowledge (cf. the 'interface issue'; see Han & Finneran, 2014, for a recent review), the development of both types of knowledge should be considered in FLT by means of form-focused instruction (FFI) and meaning-focused instruction (MFI). Indeed, if FFI did not have any positive effects on the attainment of a high degree of proficiency in a given foreign language (L2), FLT should be seriously questioned. However, it should be observed that the borders between options of MFI and FFI are not always clear-cut, which affects the coding of instructional treatments in empirical studies and meta-analyses as well as theoretical discussions on this issue.

The purpose of this study is to present the validation of a scale – by way of psychometric techniques – that enables the objective and reliable measurement of the degree of FFI and MFI loads of FLT textbook activities.

The rest of this article is structured in five more sections. The next one (i.e. background) explains the key terms of the study, specifies its rationale and accordingly establishes its objectives in detail. The third section (i.e. method) includes the type or research design followed, a description of the sample or the activities analysed, the participants (the judges involved in the application of the scale), the scale or the measuring instrument (description of its features, format and functioning), the data collection procedure and the data analysis (sub-sections 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 respectively). The fourth section reports the results, which are discussed in section 5 and where certain lines of future research are suggested too. Several concluding remarks are included afterwards.

2. BACKGROUND

Both MFI and FFI have been extensively defined in the specialised literature. MFI treats the L2 as a tool for communication and not as an object of study, that is, it is based on and triggers a holistic use of language (Samuda & Bygate, 2008; Loewen, 2011; Long, 2015). FFI, following Ellis' comprehensive definition (2001: 1-2), is "any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form". Ellis (2001) classifies FFI as Focus on Forms/FonFS (treating forms in isolation as an object of study, following Long's 1988 and 1991 original definition); incidental Focus on Form/FonF (a 'reactive' approach in response to a communicative need when using the language for communicating, also according to Long's 1988 and 1991 original conceptualisation) and planned Focus on Form/FonF, which refers to Spada's (1997) 'proactive' attention to preselected language items during communicative activities (instead of Long's exclusive reactive option). Later, Ellis (2016) specified that Focus-on-Form activities can be interactive or not, they can involve reception and production and can occur before or while a communicative task is being performed.

In the specialised literature (Norris & Ortega, 2000; De Graaff & Housen, 2009; Ellis, 2012, 2016; Collins, 2013; Ellis & Shintani, 2014; Goo et al., 2015; Long, 2015; among others), both 'explicit' and 'implicit' instruction are used to refer to FFI, which is primarily

aimed at developing explicit knowledge –regardless of the mode of learning, explicit or implicit, that is, with or without awareness (Ellis, 2012). Accordingly, explicit and implicit instruction are not generally correlated or mapped with explicit and implicit knowledge respectively. Due to space constraints, this specific issue will not be further dealt with in this article. For a very judicious explanation about the inaccuracy of equating implicit FFI with Focus on Form and explicit FFI with Focus on Forms, see Ellis (2012).

It should be observed that the borders between options of MFI and FFI are not always clear-cut. This circumstance is especially relevant for the coding of instructional treatments (independent variables) and instructional outcomes or measures (dependent variables) in meta-analyses which have examined the effectiveness of various general types of L2 instruction: Norris and Ortega (2000) and Goo et al. (2015) (for excellent critiques of the former study, see Ellis (2012) and Long (2015); see the second author for an acute appraisal of Goo et al.'s study). For example, in both meta-analyses, the instructional outcomes or dependent variables were referred to as 'metalinguistic judgement', 'selected responses', 'constrained constructed responses' and 'free-production' measures. The difficulties in the adequate coding of the free-production measures was overtly acknowledged by Goo et al. (2015: 470):

> Out of the nine free outcome measures analyzed, five of them (55.5 per cent) were writing tasks that might have allowed use of L2 knowledge that is not available in more spontaneous communication. In addition, instructions to participants did not always emphasize meaningful communication as the goal, but, instead, explicitly asked participants "to write carefully and correctly" (Andringa et al., 2011: 886), "to use the two past tenses, PC and IMP, with a minimum of 10 verbs" (Ayoun, 2001: 233), or "to describe a situation in the picture with a past tense form" (Kang, 2010: 590).

Goo et al. (2015) reasonably conclude that such features could have favoured explicit instruction (it should be noted that in both meta-analyses free-production measures were those more likely to tap implicit knowledge as opposed to the three other outcome measures coded). They propose to distinguish between oral and written free-production measures in an attempt to erradicate this methodological problem, which does not constitute a full solution as oral free-production measures can also include the above-mentioned instructions. Furthermore, it is obvious that the label 'free-production measures' inevitably evokes MFI and that the activities mentioned by Goo et al. clearly reveal a mixed case between MFI and (explicit) FFI, that is, a continuum between MFI and FFI: the students were free concerning the content they had to write about but not regarding all the forms used to express such content.

Likewise, the problem with an absolute, non-graded classification of instructional options as either pertaining to FFI or MFI is very visible in the case of consciousness-raising (C-R) activities (Loewen, 2011; Ellis, 2012) and dictoglosses (Loewen, 2011). Regarding the former, Doughty and Williams (1998) and Norris and Ortega (2000) classified them as a type of FonF (explicit). However, they "do not appear to conform to a strict definition of FonF by having a primary emphasis on meaning" (Loewen, 2011: 588), given that their goal is to inductively infer how a language form works (that is, explicit knowledge) and thus Ellis (2012: 226) noted that they could be considered a type of FonFS (also explicit). Moreover, Ellis (2012: 226) argues that C-R activities can also be classified as 'tasks' (in the terms of Task-based Language Teaching), since "the learners are required to talk meaningfully about a language point using their own resources" when trying to induce the rules underlying such a language point. Long (2015) adopts a solution for the classification of C-R activities as based on the timing of their provision: he argues that, when used proactively, they belong to FonFS, while, when used reactively within a communicative lesson, they pertain to FonF. This

solution, though attractively pragmatic, would not be applicable to the analysis of textbook activities prior to their implementation in real classes – precisely due to the reactive nature of FonF as favoured by Long (1988, 1991, 2015, among others). Concerning dictoglosses, a very popular type of activity used in collaborative writing (Swain & Lapkin, 2001; Lapkin, Swain & Smith, 2002), they have been classified as a FonF activity (Doughty & Williams, 1998; Loewen, 2011): while attempting to reconstruct the text previously listened to, students may notice a gap in their interlanguages which impede them from being accurate in their transmission of meanings. However, the primary goal of a dictogloss is the reconstruction or transcription of a language text, that is, a meaning-focused activity (MFI). Both C-R activities and dictoglosses seem to confirm the existence of a gradation in FFI and MFI – which thus supports the notion of a continuum between FFI and MFI.

In the light of the above-mentioned research methodological and theoretical issues and in the same way that efforts have been been made to operationalise measures of explicit and implicit knowledge in Second Language Acquisition (SLA) (Ellis, 2005; Erlam, 2006; Bowles, 2011; Kim & Nam, 2016), this article is aimed at reporting the results of the validation of a scale. This scale is targeted at measuring the degree of FFI and MFI loads of FLT textbooks, both objectively and reliably. To my knowledge, there is only an initial version of a similar scale in Criado, Sánchez and Cantos (2010), which was considerably modified as will be explained in sub-section 3.4.

More specifically, this article pursues two objectives. The first one consists in studying the assessment of the concurrent criterion validity of the scale scores as granted by 5 English as a Foreign Language (EFL) secondary-school teachers and 5 applied linguists on 15 randomly-selected EFL textbook activities. Criterion validity assesses whether a test reflects a set of abilities in a current (concurrent validity) setting as measured by a criterion or an outside measure used to establish validity. Both measures (the test and the criterion) take place within the same period (Salkind, 2012). Concurrent criterion validity is established by means of correlating the scores of the instrument with those of the criterion. In our case, the test is the scale, the set of abilities correspond to the degree of FFI and MFI loads of textbook activities, the external criterion is the judges' holistic assessment of the degree of FFI and MFI loads of the textbooks' activities and the scale scores are correlated with the holistic scores. The second objective is to study the inter-rater reliability coefficients of both types of assessments (holistic and scale-based).

Three relevant considerations should be taken into account. Firstly, this scale is meant to be applied to activities as these are considered the basic operational unit in Language Teaching by which textbooks and teachers implement instruction in real classrooms (Swaffer, Arens & Morgan, 1982). Secondly, textbook activities supply the material on which this scale is constructed and they also constitute the material to which the scale is applied, given that textbooks are a fundamental element of the teacher's kit – if not the most important one (Tomlinson, 2012; Guerrettaz & Johnson, 2013; Harwood, 2014). Thirdly, this scale focuses on the external sources of learning – textbook activities in this case – and it is not meant to measure or to extract direct conclusions regarding certain students' related variables such as affective factors, strategies of learning, learning processes (explicit and implicit learning), the product of learning (explicit or implicit knowledge) or the relationship between these four variables.

3. METHOD

3.1 *Research design*

This study used a repeated-measure design. The same subjects (judges) had to assess the activities holistically and with the scale in one single moment in a sequential way so as to ensure that the extraneous variable 'time' did not interfere in the assessments that they had to perform. Accordingly, the judges were asked to undertake the holistic assessment before the scale assessment to prevent the latter from exerting any kind of influence on the former in such a way that their initial impressions about the FFI and MFI loads were correctly recorded.

3.2 *Sample: activities*

The sample of this study was constituted by the textbooks' activities to which the scale was applied. 10 textbooks were selected for this purpose. They belonged to the major publishing houses in the international market (4 to Cambridge University Press, 3 to Oxford University Press, 3 to Pearson Longman) and their dates of publication ranged from 2004 to 2013. It should be taken into account that this work was not aimed at performing a cluster analysis to study whether there existed differences in the judges' scale analyses due to the belonging of the activities to different publishing houses. Moreover, since the unit of the analysis in this study is the 'activity', activities were selected in an attempt to cover all the possible activity populations independently from the textbooks to which they belonged.

3 textbooks were targeted at an A2 level (elementary); 2 at a B1 level (intermediate); 3 at a B2 level (upper-intermediate); and 2 at a C1 level (advanced). Three observations should be considered:

a) No textbooks were selected from the A1 level as it is embedded by the A2 level.
b) The B1 and B2 levels were grouped together as a single level for the purposes of using the three traditional linguistic levels, which respectively correspond to the *Common European Framework*'s (2001) distinction between Basic User, Independent User and Proficient User.
c) The C2 level does not exist in any of the textbook series from which the activities were chosen.

The second (middle) third of each textbook was randomly selected by the research as the part from which to extract the activities. All the activities from the second third of each textbook were compiled. Three lists of all the activities within each respective level were devised. In each list, all the activities were consecutively numbered and randomly selected on the basis of the numbers that they had been assigned. 5 activities from each one of the three levels were selected, thus, totalling 15 activities.

3.3 *Participants: judges*

10 judges analysed the activities holistically and by means of the scale. There were two groups of judges: 5 applied linguists who were specialists in FLT and teacher training, and 5 EFL secondary-school teachers. The 10 judges were Spanish-native speakers. In both groups there was 1 man and 4 women. All the members of the applied linguists' group were qualified as to a PhD level, their mean age was 53 years and their professional experience ranged from 8 to 33 years. As to the secondary school teachers' group, 2 of them had earned their PhDs; their mean age was 41 years and their professional experience ranged from 11 to 23 years.

3.4 *Measuring instrument: the scale*

As explained in section 2, there exists a FFI-MFI continuum since (textbook) activities do not always reflect a clear and neat ascription to either FFI or MFI. Thus, in terms of their belonging to FFI and MFI, activities should not be measured with absolute global values of YES or NO. Accordingly, 8 features were distinguished to represent the 'instruction' variable. The 8 selected features are the following ones; as can be seen, they are worded so that they cover FFI and their opposites refer to MFI:

1. Involves the use of metalanguage.
   Metalanguage, as defined by Ellis and Shintani (2014: 341) is "the terminology available for analyzing or describing a language. It can be highly technical (e.g., 'hypothetical conditional') or everyday (e.g., verb)". It entails the most explicit approach of FFI. Metalanguage is inherently related to the formal intricacies of a language and underlies metacognition about language or metalinguistic awareness (DeKeyser, 2009), which I have purposefully omitted, as cognition and awareness reflect learners' internal processing of language.
2. Involves analysing language.
   This feature covers whether an activity pushes the students to analyse language (examining how a formal feature works) without overtly making them resort to using metalanguage. It can be considered to represent the second most distinctive characteristic layer of the explicit approach to FFI. It should be considered that analysing language does not necessarily imply that one uses (or even masters) metalanguage; in fact, a person can assess whether a sentence is grammatically correct or not without using any metalanguage (Berry, 2005, 2014).
3. Favours formal accuracy.
   Favouring formal accuracy can be the purpose of an activity regardless of whether it involves the use of metalanguage or analysing language. 'Accuracy' here is employed from the point of view of the pedagogical target of the activity, which emphasises specific form/s; as such, the activity implies precision in the language presentation and manipulation. Accordingly, formal accuracy is not only favoured by explicit deductive and inductive presentations (direct and indirect in terms of Ellis & Shintani, 2014) but also by implicit inductive presentations as in the Audiolingual Method and the Situational Language Teaching Method (Howatt with Widdowson, 2004; Sánchez, 2009). In these Structural Methods, learners engaged in language practice by listening to input-enriched texts and reciting them (without any overt direction to pay attention to the forms). Interestingly, by developing linguistic habits in an implicit inductive way, these methods did not enable learners to become communicatively competent as, among other reasons, most of their instructional techniques and activities were form-focused – and thus, accuracy-targeted (DeKeyser, 1998; Criado, 2010).
4. Aims at the controlled input- or output-based use of the language.
   Controlled use implies that the learners do not have any leeway for the selection of the forms that they have to process or use, as opposed to 'free' use. Controlled use of language is one possible purpose of 'practice' in language pedagogy, a wide-ranging concept. It is defined by DeKeyser (2007: 1) as "specific activities in the second language, engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language". The systematic engagement is intrinsically related to a fundamental characteristic of practice: repetition (DeKeyser, 2007). Output-driven repetition is fostered in production-based instruction, in which drills are a classical type of activity (Shintani, Li & Ellis, 2013; Ellis & Shintani, 2014). Other types of production-based activities which imply controlled use of language may be distinguished; for

example, focused output-based tasks (Ellis, 2003, 2012) or structure-based production tasks designed on the principle of 'task usefulness' (Loschky & Bley-Vroman, 1993). There also exists input-driven repetition, which is encompassed in comprehension-based instruction – such as input enrichment, processing instruction and input-based focused tasks (Ellis, 2012).

5. Entails mechanical repetition.
   In a landmark article, Paulston (1970) distinguished between mechanical drills, meaningful drills and communicative drills. Mechanical drills are those which just require the learners to repeat the targeted structure without any recourse or attention to meaning or the pragmatic and social context in which they occur. For instance, from "I ate an apple. What did I eat?" to "You ate an apple" (DeKeyser, 1998: 50). Meaningful drills are those that require the students to process meaning but there is not an information gap to be covered; for example, "Is this a pen or a pencil? It's a pencil" (DeKeyser, 1998: 50). Communicative drills are those that entail filling an information gap while using specifically targeted language forms; for example, when asking "what would you do if you saw your best friend cheating in an exam?", possible responses could be "I would do nothing" or "I would tell the teacher" among others (Criado, 2010: 51). From the point of view of content, the answers are unpredictable as they depend on the students' impressions and feelings in the specific situation depicted by the question; from the perspective of forms, the students are induced to use the targeted language feature (second conditional in this case).
6. Involves the use of L1.
   L1 use is arguably more relevant in a foreign language context than in a second language setting given the former's meager qualitative and quantitative characteristics of L2 input. Likewise, L1 use seems to be more relevant in a monolingual foreign language context than in a multilingual foreign language setting. Specifically, this feature measures whether the activity instructions encourage students to make cross-linguistic comparisons in terms of grammar, vocabulary, spelling or pronunciation; in other words, whether learners are induced to resort to their L1 formal linguistic system for the accomplishment of the activity (translation being one of such uses). Cross-linguistic comparisons enable learners to become aware of formal features that might be ignored otherwise (James, 1980; Cummins, 2007) and constitute a case of positive transfer. They can also entail exercising students' analytical and reasoning skills.
7. Mostly triggers individual work on the language.
   In traditional FLT, individual work has been ascribed to form-focused activities and to reading, writing and listening activities. However, recent cognitive-interactionist and socio-cultural theories in SLA consider interaction as having both a knowledge-getting and a skill-getting function, though determined in different ways (Ellis & Shintani, 2014). Accordingly, verbal interaction can facilitate the learning of new language forms in activities not primarily targeted at practising the speaking skill – for instance, in C-R activities. The fact that an activity has a formal focus does not automatically entail that it is intended to be performed individually and/or that the learners will use language in a non-communicative way (see feature 8). Likewise, more and more textbooks are incorporating aural- and written-text activities where pair- and group-work dynamics are overtly stated in the instructions in order to encourage learners to interact.
8. Requires using the language for a non-communicative purpose.
   As opposed to the previous structural and humanistic methods, whose real outcome was the correct accurate reproduction of sentences, the intended goal of communicative approaches (for instance, Task-based Language Teaching and Content and Language Integrated Learning) consists of being able to use the language for communicative

purposes (Howatt with Widdowson, 2004; Richards & Rodgers, 2014; among others). Authentic communicative purposes entail filling an information, reasoning or opinion gap (Prabhu, 1987) which motivates learners to communicate receptively and productively in both oral and written modes; in other words, processing and using the language for meaning and not for formal intricacies.

As stated in section 2, there exists an initial version of the scale presented in this article (Criado, Sánchez & Cantos, 2010), which was extensively adapted and refined as follows:

a) Their scale was designed to measure explicit and implicit teaching. Such labels were changed to FFI and MFI respectively to better capture the fact that FFI can be explicit or implicit (as explained in section 2).
b) Their scale used 'goals' and the 'pedagogical strategies to attain such goals' as parameters of analysis to be taken into account when analysing the activities with the scale. It became evident after the application of that scale that the consideration of both parameters, though conceptually compelling, excessively complicated its application (usability being one the characteristics of the prospective scale to be achieved).
c) The new scale includes eight features instead of nine. Out of the nine features of the previous scale, the first and the third ones ('looks for awareness of formal aspects' and 'promotes declarative knowledge of the language') were removed as they covered students' internal learning processes and types of knowledge. The second feature ('focuses on metalinguistic information of the language') was split up in two different features respectively covering the use of metalanguage and analysing language. The fourth feature ('focuses primarily on form') was removed as it was considered not to be sufficiently discriminatory regarding the fifth one ('favours linguistic accuracy' which, in turn, was rephrased as 'favours formal accuracy'). The sixth feature ('uses non-authentic materials') was replaced by 'requires using the language for a non-communicative purpose', which was believed to be more pertinent for the continuum between FFI and MFI: what the activity entails doing, that is, manipulating language for non-communicative purposes, rather than the authenticity of the materials. There can be activities which ask the learners to perform a communicative action with an authentic text – reading for pleasure, for instance – and activities which require the learners to perform a more linguistic-based action with an authentic text too – such as inferring the unknown meaning from co-text and context or detecting anaphoric or cataphoric referents. The eighth feature ('aims at controlled use of the language') was extended so that it comprehensively covered both input- and output-based language. A new feature (number seven: 'mostly triggers individual work on the language') was added since it was considered to include an important aspect of the 'instruction' variable: classroom dynamics (individual versus non-individual work). The ninth feature was maintained without the adjective 'non-meaningful' ('entails mechanical repetition'), which was believed to be redundant.
d) Finally, a modest attempt to validate that scale was implemented. It consisted of computing the correlation indices of the scale scores on 23 textbook activities provided by three expert judges (no statistically significant differences were found).

In the same way as in Criado, Sánchez & Cantos (2010), a Microsoft Excel spreadsheet file was designed in order to host the scale and to allow for the automatic computing of the FFI and MFI values. Table 1 offers a sample analysis of an authentic textbook activity with the scale.

As can be seen, the instructions of the activity do not ask the students to use any metalanguage or to analyse language. Still, the activity implies formal accuracy, since pedagogically speaking it has a clear and distinct formal target, the past simple tense, which the students must practise in an output-based way by creating and using sentences which contain instances of this tense. However, such practice is not mechanical given that the activity is an

example of a communicative drill (see above). Also, it does not trigger the use of the students' L1 (explicitly, at least) and requires interacting with the whole class. The language is used for communicative purposes (finding a person who did several things and obtaining more related information) or as communicative as possible, taking into account the targeted level and the limitations of the FL classroom (which can never be fully equated with the outside world and thus it always contains a certain degree of formal focus –however subliminal).

The FFI features correspond to the Yes value, as opposed to the MFI column, whose value is No in all its cells. For each feature in each activity the '1' number should be typed in the FFI or MFI cell as appropriate; nothing is typed in the other cell. Data are computed in the spreadsheet file as follows: the points in each column are counted, the result being 8 points overall between the two columns (FFI and MFI), as can be seen in the 'Total score (out of 8 points)' row. Then that total is computed to 10 points. This computing appears in the 'Total score (out of 10 points)' row (see Figure 1).

Table 1: ***The scale for measuring the FFI and MFI loads of language teaching activities (sample analysis)***

New English File Elementary (2004). File 4B. Page 55 Stand up and move around the class. Ask *Did you…?* questions. When somebody answers *Yes, I did* write down their name and ask the next question, e.g. *Where to?* *Find a person who…* *Travelled by plane last year. _____Where to?* *[Example provided in the textbook]: Did you travel by plane last year? Yes, I did. Where to?* *Started learning English a long time ago. _____ When?* *Etc.*	***FFI*** ***Yes***	***MFI*** ***No***
1. Involves the use of metalanguage		1
2. Involves analysing language		1
3. Favours formal accuracy	1	
4. Aims at the controlled input- or output-based use of the language	1	
5. Entails mechanical repetition		1
6. Involves L1 use		1
7. Involves individual work on language		1
8. Requires using the language for non-communicative purposes		1
TOTAL score (out of 8 points)	2	6
TOTAL score (out of 10 points)	**2.50**	**7.50**

As can be observed, the scale is inclusive and allows for different weights of FFI and MFI, in such a way that the higher the FFI load, the lower the MFI load and viceversa. This is linked to the FFI-MFI continuum explained above. Accordingly, the scale reflects just one variable –instruction– in accordance with its unidimensional nature. This 'instruction'variable has two end-points: 10 (maximum) and 0 (minimum). The continuous nature of the 'instruction' variable is depicted in Figure 1 (which follows on from the preceding example in Table 1):

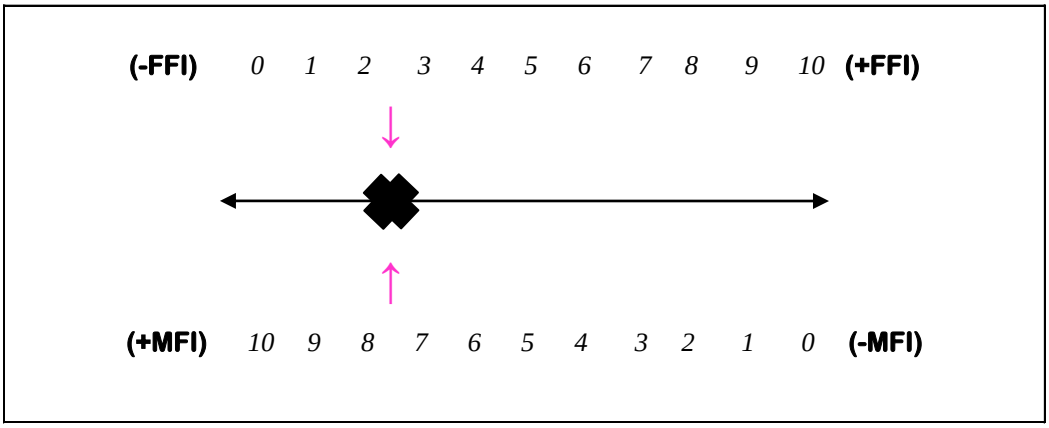


Figure 1: ***The continuum of the 'instruction' variable (example derived from Table 1)***

3.5 *Data collection*

The two groups of judges were contacted by e-mail by the researcher. After they consented to participate in the study, each one of them received an e-mail with these two following sets of material:

a) A booklet containing the instructions of the tasks they had to perform and the 15 scanned activities. The instructions offered a brief explanation of the purpose and rationale for the research, a description of the scale and its features, the guidelines to fill in the two spreadsheets provided and a sample analysis of both the holistic and the scale-based assessments. As to the activities, they were randomly ordered and there was not any reference either to the title of the textbooks to which each activity belonged or to the textbooks' targeted level. The transcripts of the listening activities included were also incorporated.
b) Two spreadsheet files: 'HOLISTIC SCORES_Judge n' and 'SCALE SCORES_Judge n'. As stated in sub-section 3.1, in accordance with the above-mentioned research design of the study (repeated-measure design), the judges were instructed to analyse the activities holistically before applying the scale so as to ensure that their holistic assessments were not influenced by their scale-based assessments.

The first spreadsheet file included three cells in the first row –'activity number', 'FFI holistic score' and 'MFI holistic score'– followed by fifteen rows for each one of the 15 activities labelled and consecutively numbered. The judges were asked to assign a holistic score from 0 to 10 to both the loads of FFI and MFI according to their judgement in all the 15 activities so that that the addition of both scores was 10 (for example, in the sample activity of Table 1: 3 FFI scores + 7 MFI scores = 10 scores).

The second spreadsheet file contained fifteen identical scale templates with which the judges were asked to analyse the activities. For the purposes of explaining the functioning of the spreadsheet file, the guidelines of the booklet included similar indications to those included in sub-section 3.4.

Overall, the researcher's initial contact with the judges and the collection of all the data took place between January and March 2015.

3.6 *Data analysis*

After their collection, the data were submitted to a database file of the *Statistical Package for the Social Sciences (SPSS)*, version 21. The FFI holistic scores, the MFI holistic scores, the FFI scale scores and the MFI scale scores granted by each one of the 10 judges for each one of the 15 activities were coded and tabulated.

It should be observed that the MFI scores were also recorded in the SPSS file to have them readily available although they were not necessary for the computation of the concurrent criterion validity. As indicated in sub-section 3.4, the scale measured a single variable –instruction– which was considered to be unidimensional (not multidimensional).

Regarding concurrent criterion validity, Pearson correlation coefficients were computed between the holistic scores on FFI (external criterion) and the scores on FFI resulting from the application of the scale by every single judge in all of the 15 activities, as well as by the language level of the activities.

The inter-rater reliability from the application of the scale was studied by means of the intraclass correlation coefficient (ICC) in the Mixed Two-way category (Absolute Agreement) of SPSS. ICC was computed overall, by the groups of judges, by the language level of the activities and by these two factors together.

4. RESULTS

This section reports the results for the concurrent criterion validity of the scale and the inter-rater reliability.

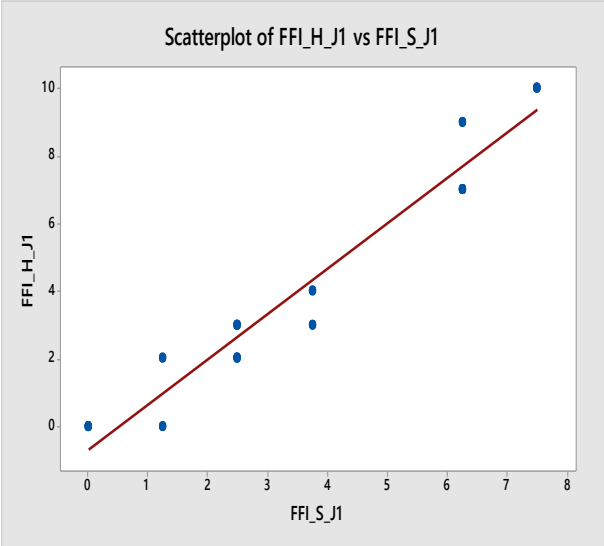
4.1 *Concurrent criterion validity*

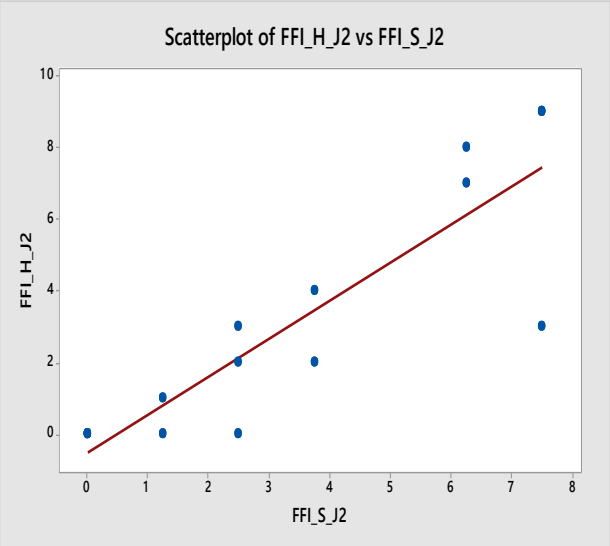
As stated in the previous section, the judges' scores from the application of the scale were correlated with the judges' holistic scores (external criterion) as a measure of concurrent criterion validity by means of Pearson correlation coefficients. Table 2 shows the results of the correlations between the scale scores and the holistic scores for each judge in all the 15 activities as well as by the language level of the activities.

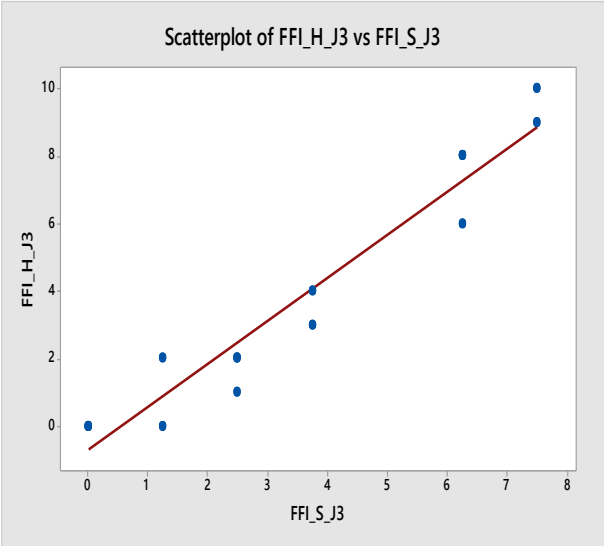
Table 2: ***Pearson criterion validity coefficients***

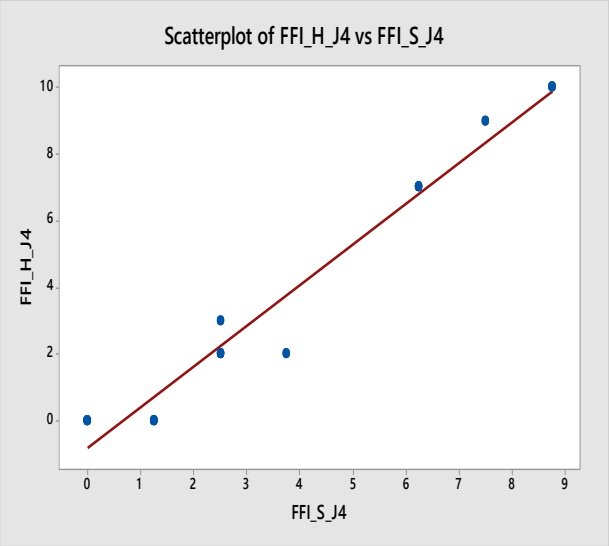
	Total	Elementary Level (A2)	Intermediate Level (B1-B2)	Advanced Level (C1)
Applied Linguists	r_{xy}	r_{xy}	r_{xy}	r_{xy}
Judge 1	0.973	0.974	0.984	0.955
Judge 2	0.876	0.996	0.899	0.942
Judge 3	0.968	0.994	0.972	0.959
Judge 4	0.982	0.990	0.973	0.989
Judge 5	0.962	0.940	0.987	0.986
Teachers	r_{xy}	r_{xy}	r_{xy}	r_{xy}
Judge 6	0.986	0.990	0.977	0.995
Judge 7	0.939	0.899	0.984	0.948
Judge 8	0.991	0.990	0.994	0.995
Judge 9	0.955	0.930	0.987	0.996
Judge 10	0.928	0.889	0.980	0.932

As can be seen in the applied linguists' group, the concurrent criterion validity coefficient ranged between 0.876 and 0.982, while in the group of the EFL secondary school teachers, the concurrent criterion validity value ranged between 0.928 and 0.991. Consequently, there is very satisfactory concurrent criterion validity regardless of both the judge and the language level of the activities, since all the estimated coefficients are above 0.70. The scatterplots for the coefficients of each judge in all the 15 activities (see Figure 2 below) show that, overall, the data points tightly fit the regression line, thus indicating that the scale has very satisfactory concurrent criterion validity.









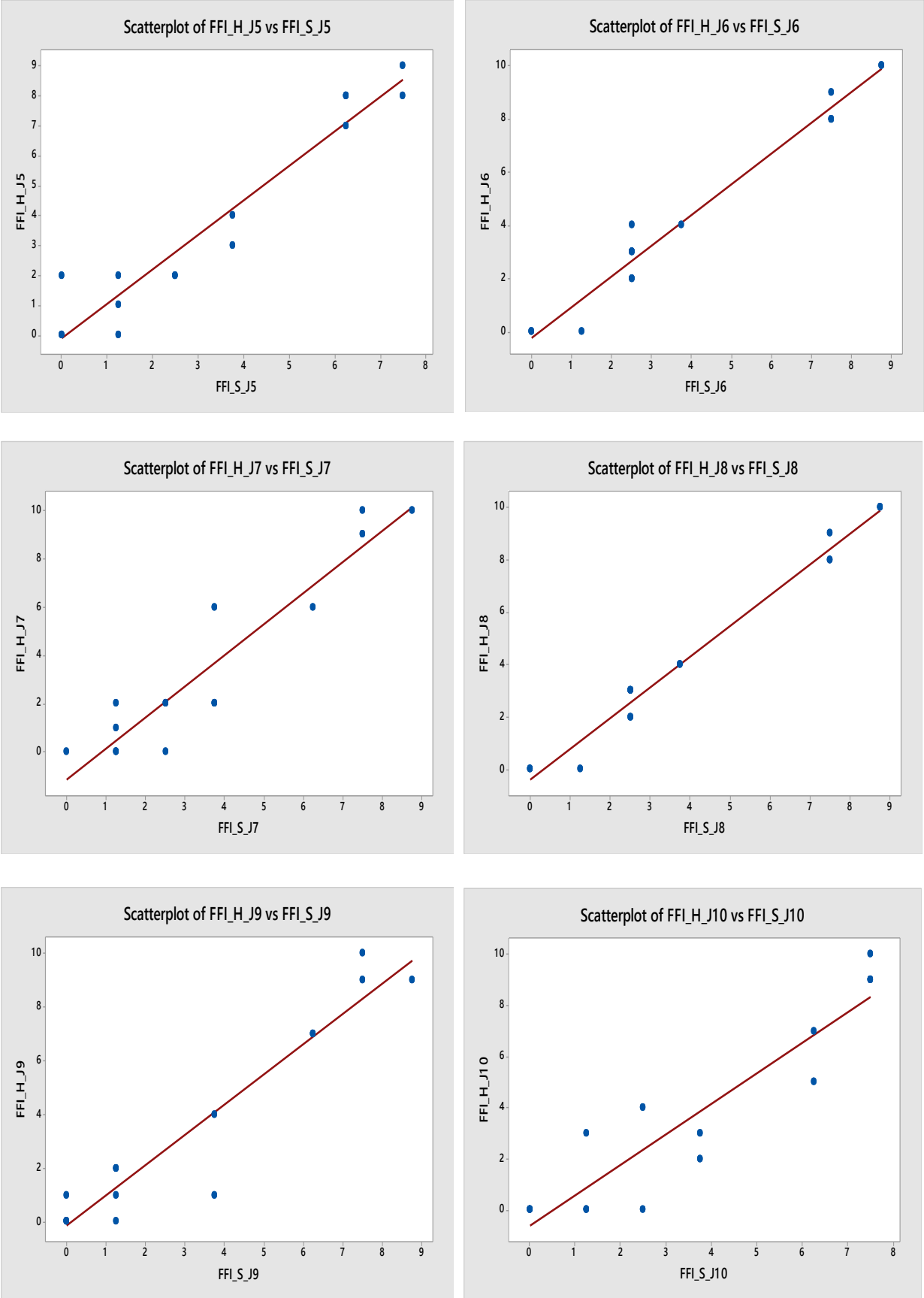


Figure 2: ***Scatterplots of the FFI assessment – holistic and scale-based – by each judge in all the 15 activities (H = holistic score; S = scale score; J = judge)***

4.2 *Inter-rater reliability*

As indicated above, the intraclass correlation coefficient (ICC) in the Mixed Two-Way (Absolute Agreement) category of SPSS was computed to obtain:

a) the overall inter-rater agreement between the scale scores and the holistic scores regardless of groups of judges and the language level of the activities;
b) the inter-rater agreement by both groups of judges and by the language level of the activities separately;
c) the inter-rater agreement by groups of judges and by the language level of the activities jointly.

4.2.1 *Overall ICC*

The ratio of the overall agreement for the scale scores, regardless of the groups of judges (applied linguists vs. EFL secondary school teachers) and the language level of the activities (elementary, intermediate and advanced), yielded a value of 0.942. This indicates a high level of agreement between judges when assessing the FFI and MFI load of the textbooks' activities. The inter-rater agreement of the holistic assessment was also high, with a value of 0.920, and thus it is not statistically different from that of the scale assessment.

Table 3: ***Overall ICC***

	ICC CI95%[1]
Scale analysis	0.942 0.890-0.976
Holistic analysis	0.920 0.853-0.967
p	0.882

[1] CI95%: confidence interval 95% for ICC

4.2.2 *ICC by groups of judges*

Table 4 below offers the results of ICC by groups for both types of assessments. In relation to the scale scores, the applied linguists' group obtained an ICC of 0.958, compared to the 0.936 ICC of the EFL secondary school teachers' group. In both cases, the coefficient of inter-rater agreement between groups of judges is excellent, with no statistically significant differences between the two groups of judges ($p > 0.05$). If we compare these ICC values of the scale assessment with those obtained in the holistic assessment in each group of judges, the applied linguists obtained an ICC of 0.909 for the holistic assessment, which statistically differs from that obtained through the scale. Regarding the EFL secondary school teachers' group, the ICC value obtained for the holistic assessment was 0.936, which is statistically equivalent to that obtained with the scale.

Table 4: ***ICC by groups of judges***

	Applied Linguists ICC CI95%[2]	Teachers ICC CI95%	p
Scale scores	0.958 0.916-0.984	0.936 0.868-0.975	0.878
Holistic scores	0.909 0.824-0.964	0.936 0.864-0.975	0.862
p	0.047	1.000	

[2] CI95%: confidence interval 95% for ICC

4.2.3 *ICC by the language level of the activities*

ICC values by the language level of the activities are shown in Table 5. There are not any statistically significant differences between the scale and the holistic assessments in each level independently considered. The ICC values for the scale and the holistic analyses between the three levels jointly considered revealed no statistically significant differences either. All the ICC values obtained are above 0.90, which involves a high rate of inter-rater agreement.

Table 5: ***ICC by the language level of the activities***

	Scale ICC CI95%[3]	Holistic ICC CI95%	p
Elementary level (A2)	0.941 0.835-0.993	0.938 0.827-0.992	0.974
Intermediate level (B1-B2)	0.962 0.884-0.995	0.905 0.750-0.988	0.781
Advanced level (C1)	0.943 0.839-0.993	0.941 0.836-0.993	0.991
p	0.618	0.586	

[3] CI95%: confidence interval 95% for ICC

4.2.4 *ICC by both groups of judges and the language level of the activities*

ICC values show a high rate of inter-rater agreement regardless of the group of judges and the language level of the activities for both the scale and the holistic assessments (see Table 6 for all the results).

Table 6: ***ICC by both groups of judges and the language level of the activities***

	Applied Linguists		Teachers		
	Scale ICC CI95%[4]	Holistic ICC CI95%	Scale ICC CI95%	Holistic ICC CI95%	p
Elementary level (A2)	0.955 0.854-0.995	0.947 0.831-0.994	0.918 0.743-0.990	0.935 0.781-0.992	0.859
Intermediate level (B1-B2)	0.976 0.913-0.997	0.857 0.611-0.981	0.958 0.849-0.995	0.962 0.869-0.995	0.532
Advanced level (C1)	0.957 0.859-0.995	0.942 0.816-0.993	0.956 0.859-0.995	0.933 0.792-0.992	0.916
p	0.832	0.072	0.213	0.547	

[4] CI95%: confidence interval 95% for ICC

5. DISCUSSION

As observed in Table 2, the Pearson correlations show that the scale has very satisfactory concurrent criterion validity, regardless of each individual judge and of the language level of the activities. This result is very promising although it would be convenient to check whether this tendency of high concurrent criterion validity coefficients would be solidly maintained regarding both single judges and different language levels of the activities with a large sample of activities to be assessed.

Concerning ICC, there is very high overall inter-rater reliability as shown in Table 3. Excellent inter-rater agreement coefficients for both the scale and the holistic assessments were also yielded in both groups of judges (Table 4). There was a statistically significant difference between the scale assessments and the holistic assessments in the group of the applied linguists –the ICC value of the scale assessment being higher than that of the holistic assessment. It could be argued that this statistical difference is due to the fact that the scale is an instrument that compiles all the criteria that the applied linguists contemplate as derived from their academic knowledge, which they always have in mind when they tackle this issue. Interestingly, the degree of agreement of the EFL secondary school teachers was similar in both types of assessments.

An excellent inter-rater agreement coefficient was also attained when ICC was computed by the language level of the activities (Table 5) and by the two factors jointly considered (Table 6), with no statistically significant differences in either case. In the intermediate level for the applied linguists' group there seems to be a tendency by which the degree of agreement in the holistic assessment is lower than that of the scale assessment, as opposed to the remaining ICC values. However, this difference could not be statistically verified, which is probably due to the size of the sample (5 activities in each language level).

Overall, it is noteworthy that the inter-rater reliability results of the scale assessment are high and do not appear to be influenced by the membership of the judges to one or another group (Table 4), or by the language level of the activities (Table 5) or by both factors jointly considered (Table 6). It seems that the features selected to design the scale were correctly selected and accurately reflect the characteristics of FFI and MFI given that the results show a high agreement in the assessment of the FFI and MFI loads in the textbooks' activities. Thus, despite the small sample of activities assessed, all the results suggest that the value of the scale lies in that it constitutes a tool that can systematise and facilitate the assessment of the FFI and MFI loads of textbook activities performed by both applied linguists and EFL secondary school teachers.

Even though the results of this study are positive and encouraging, several lines of future research into the concurrent criterion validity of the scale as well as the inter-rater agreement of the judges can be indicated.

Firstly, as stated above, augmenting the sample of the activities to be assessed would shed light on whether the tendencies of the results obtained in this study are maintained or not. More specifically, it would be very convenient to verify whether the concurrent criterion validity coefficients would be satisfactory or not when computed by the language level of the activities. Likewise, studying judge-group assessments in the ICC by the language level of the activities would contribute to ascertaining whether the ascription to a specific group of judges involves different patterns in the assessment of the FFI and MFI loads of the activities and whether such different patterns are kept constant across different language levels of the activities or not.

Secondly, it would be interesting to transform the variable of the responses of the scale items from dichotomic to ordinal (as in a Likert scale) and to study whether this would entail any changes in the assessments of the FFI and MFI loads of textbook activities by different groups of judges.

Thirdly, a cluster analysis of the activities would contribute to discerning whether different patterns of FFI and MFI loads could be distinguished in the activities according to their language level.

Fourthly, a cluster analysis of the judges would contribute very important data to verify whether different types of groups of judges constitute clearly distinct groups of judges in terms of their assessments. Such groups of judges could be composed of not only applied linguists and teachers with varied professional profiles, but also teacher trainees, since the

quantity –and quality– of years of professional experience is a factor that could potentially influence the differences in the judges' assessments.

6. CONCLUSION

The objective of this study was to investigate the concurrent criterion validity of a scale designed to measure the FFI and MFI loads of FLT textbook activities. For that purpose, the holistic and scale scores on 15 EFL textbook activities as granted by two groups of judges – applied linguists and EFL secondary school teachers – were analysed. High concurrent criterion validity coefficients with no statistically significant differences were obtained as computed overall and by the language level of the activities. Excellent inter-rater agreement coefficients were shown overall, in both groups of judges (with statistically significant differences revealed in the group of the applied linguists for the scale and holistic assessments, the former being higher), by the language level of the activities and by both factors jointly considered. These results suggest that the scale constitutes a valuable and standardised tool of analysis of the FFI and MFI loads of FLT textbook activities.

Overall, I strongly believe that three types of users employing any foreign language can benefit from the scale presented here:

a) Teachers and material authors. The scale could be useful in their design and selection of the teaching activities to be implemented in textbooks and in L2 classrooms.
b) Teacher trainees. Their application of the scale would be useful to train them to differentiate between FFI, MFI and intermediate cases as well as how to identify the types of activities that foster each instructional approach.
c) Researchers. This scale would allow them to accurately and reliably measure the degree of FFI and MFI of both instructional treatments (independent variables) and outcomes (dependent variables) in (quasi-)experiments and meta-analyses aimed at studying the degree of efficacy of instruction on L2 learning. In other words, the inconsistencies indicated in section 2 as to the coding and classification of these types of variables would be erradicated.

In conclusion, I endorse the view that the design and validation of this scale constitutes an example of language-pedagogy research whose future implementation could inform not only FLT theory and practice, but also Teacher Training and SLA – with the hope of promoting a fruitful collaboration between these three disciplines.

ACKNOWLEDGEMENTS

This work was funded by the Spanish Ministry of Economy and Competitiveness (Research Project Ref. FFI2013-44979-R), which is hereby gratefully acknowledged.

REFERENCES

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research,* 16(1), 9-37. DOI: 10.1177/1362168811423339

Berry, R. (2005). Making the most of metalanguage. *Language Awareness,* 14(1), 3-20. DOI: http://dx.doi.org/10.1080/09658410508668817

Berry, R. (2014). Investigating language awareness. The role of terminology. In A.Łyda & K. Szczesniak (Eds.), *Awareness in Action. The Role of Consciousness in Second Language Acquisition* (pp. 21-34). Bern: Springer.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: what can heritage language learners contribute? *Studies in Second Language Acquisition*, 33(2), 247-271. DOI: http://dx.doi.org/10.1017/S0272263110000756

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment.* Cambridge: Cambridge University Press.

Collins, L. (2013). Form-focused instruction. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-8). Oxford: Wiley-Blackwell. DOI: 10.1002/9781405198431.wbeal0432

Criado, R. (2010). *Activity Sequencing in Foreign Language Teaching Textbooks. A Cognitive and Communicative Processes-Based Perspective.* Saarbrücken, Germany: LAP Lambert Academic Publishing AG & Co. KG.

Criado, R., Sánchez, A. & Cantos, P. (2010). An attempt to elaborate a construct to measure the degree of explicitness and implicitness in ELT materials. *International Journal of English Studies (IJES)*, 10(1), 103-129. DOI: 10.6018/ijes.10.1.114001

Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics (CJAL)*, (10)2, 221-240.

De Graaff, R. & Housen, A. (2009). Investigating the effects and effectiveness of L2 instruction. In M. Long & C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 726-755). Oxford: Blackwell. DOI: 10.1002/9781444315783.ch38

DeKeyser, R. M. (1998). Beyond focus on form: cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 42-63). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2007). Introduction. Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language. Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1-18). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. Long & C. Doughty (Eds.), *Handbook of Second Language Teaching* (pp. 119-138). Oxford: Blackwell. DOI: 10.1002/9781444315783.ch8

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (2nd ed.) (pp. 94-112). New York: Routledge.

Doughty, C. & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 197-261). Cambridge: Cambridge University Press.

Ellis, R. (2001). Introduction: investigating form-focused instruction. *Language Learning*, 51(Supplement 1), 1-46. DOI: 10.1111/j.1467-1770.2001.tb00013.x

Ellis, R. (2003). *Task-based Language Learning and Teaching.* Oxford: Oxford University Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: a psychometric study. *Studies in Second Language Acquisition*, 27(2), 141-172. DOI: http://dx.doi.org/10.1017/S0272263105050096

Ellis, R. (2012). *Language Teaching Research and Language Pedagogy*. Oxford: John Wiley & Sons.

Ellis, R. (2013). Principles of instructed second language learning. In M. Celce-Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a Second or Foreign Language.* (4th ed.) (pp. 31-45). Boston: Heinle and Heinle.

Ellis, R. (2015). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 417-442). Amsterdam: John Benjamins.

Ellis, R. (2016). Anniversary article. Focus on form: A critical review. *Language Teaching Research*, 20(3), 405-428.

Ellis, R. & Shintani, N. (2014). *Exploring Language Pedagogy through Second Language Acquisition Research.* London: Routledge.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: an empirical validation study. *Applied Linguistics*, 27(3), 464-491. DOI: 10.1093/applin/aml001

Goo, J., Granena, G., Yilmaz, Y. & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 443-482). Amsterdam: John Benjamins.

Guerrettaz, A. M. & Johnston, B. (2013). Materials in the classroom ecology. *Modern Language Journal*, 97(3), 779-796. DOI: 10.1111/j.1540-4781.2013.12027.x

Han, Z.-H. & Finneran, R. (2014). Re-engaging the interface debate: strong, weak, none, or all? *International Journal of Applied Linguistics*, 24(3), 370-389. DOI: 10.1111/ijal.12034

Harwood, N. (2014). Content, consumption, and production: three levels of textbook research. In N. Harwood (Ed.), *English Language Teaching Textbooks: Content, Consumption, Production* (pp. 1-44). London: Palgrave.

Howatt, A.P.R. with Widdowson, H. (2004). *A History of English Language Teaching.* (2nd ed.). Oxford: Oxford University Press.

James, C. (1980). *Contrastive Analysis.* London: Longman.

Kim, J. E. & Nam, H. (2016). Measures of implicit knowledge revisited: processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, 1-27. DOI: 10.1017/S0272263115000510

Lapkin, S., Swain, M. & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *Modern Language Journal,* 86(4), 485-507. DOI: 10.1111/1540-4781.00157
Loewen, S. (2011). Focus on form. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning. Volume II* (pp. 576-592). New York: Routledge.

Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in Second Language Acquisition: Multiple Perspectives* (pp. 115-141). Rowley, MA: Newbury House.

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In K. de Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign Language Research in Cross-Cultural Perspective* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. (2015). *Second Language Acquisition and Task-Based Language Teaching.* Oxford: Wiley Blackwell.

Loschky, L. & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks in A Pedagogical Context: Integrating Theory and Practice* (pp. 123-167). Clevedon: Multilingual Matters.

Norris, J. & Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning,* 50(3), 417-528. DOI: 10.1111/0023-8333.00136

Paulston, C. (1970). Structural pattern drills: a classification. *Foreign Language Annals,* 4(2), 187-193. DOI: 10.1111/j.1944-9720.1970.tb02033.x

Prabhu, N. S. (1987). *Second Language Pedagogy.* Oxford: Oxford University Press.

Richards, J. C. & Rodgers, T. (2014). *Approaches and Methods in Language Teachin*g. (3rd ed.). Cambridge: Cambridge University Press.

Salkind, N. J. (2012). *100 Questions (and Answers) About Research Methods.* Los Angeles: SAGE.

Samuda, V. & Bygate, M. (2008). *Tasks in Second Language Learning.* New York: Palgrave Macmillan.

Sánchez, A. (2009). *La enseñanza de idiomas en los últimos cien años. Métodos y enfoques.* Madrid: SGEL, S.A.

Shintani, N., S. Li, & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: a meta-analysis of comparative studies. *Language Learning, 63*(2), 296-329. DOI: 10.1111/lang.12001

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30(2), 73-87. DOI: http://dx.doi.org/10.1017/S0261444800012799

Swain, M. & Lapkin, S. (2001). Focus on form through collaborative dialogue: exploring task effects. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing* (pp. 99-118). Harlow: Longman.

Swaffer, J. K., Arens, K. & Morgan, M. (1982). Teacher classroom practices: redefining method as task hierarchy. *Modern Language Journal*, 66(1), 24-33. DOI: 10.2307/327812

Tomlinson, B. (2012). Materials development for language teaching and learning. *Language Teaching*, 45(2), 143-179. DOI: http://dx.doi.org/10.1017/S0261444811000528

List of the EFL textbooks from which the sample of activities was extracted:

Cunningham, S. & Moor, P. (2013). *Cutting Edge Intermediate. Third Edition. Student's Book.* Harlow: Pearson Longman.

Cunningham, S. & Moor, P. (2013). *Cutting Edge Upper-Intermediate. Third Edition. Student's Book.* Harlow: Pearson Longman.

Cunningham, G., Bell, J. with Redston, C. (2009). *Face2Face Advanced. Student's Book.* Cambridge: Cambridge University Press.

Eales, F. & Oakes, S. (2011). *Speak Out Elementary. Student's Book*. Harlow: Pearson Longman.

Oxenden, C. & Latham-Koenig, C. (2004). *New English File Elementary. Student's Book*. Oxford: Oxford University Press.

Oxenden, C. & Latham-Koenig, C. (2008). *New English File Upper Intermediate. Student's Book*. Oxford: Oxford University Press.

Oxenden, C. & Latham-Koenig, C. (2010). *New English File Advanced. Student's Book.* Oxford: Oxford University Press.

Redston, C., & Cunningham, G. (2005). *Face2Face Elementary. Student's Book.* Cambridge: Cambridge University Press.

Redston, C. & Cunningham, G. (2006). *Face2Face Intermediate. Student's Book.* Cambridge: Cambridge University Press.

Redston, C. & Cunningham, G. (2007). *Face2Face Upper Intermediate. Student's Book.* Cambridge: Cambridge University Press.