Lang. Teach. (2009), **42:4**, 504–518 © Cambridge University Press
doi:10.1017/S0261444808005697 First published online 15 January 2009

A priori assessment of language learning tasks by practitioners

Gerard J. Westhoff Utrecht University, The Netherlands
g.j.westhoff@uu.nl

Teachers' competence to estimate the effectiveness of learning materials is important and often neglected in programmes for teacher education. In this lecture I will try to explore the possibilities of designing scaffolding instruments for a priori *assessment of language learning tasks, based on insights from SLA and cognitive psychology, more specifically connectionist theory. I will subsequently outline the development and evaluation of a 'yardstick' to judge complex, integrated, life-like tasks, such as WebQuests. The possibilities will be explored of performing in-depth* a priori *task analyses as a learning task for teachers in order to enhance their competence in making 'educated guesses' about task effectiveness. Finally, an experiment will be described to determine the reliability and validity of an instrument for in-depth analysis of language learning tasks based on the theoretical framework previously described.*

1. Introduction

Several recent developments have stimulated and justified the increased interest in the competence of teachers to estimate the learning impact of instructional materials. Firstly, new technologies such as the Internet enable teachers to explore the possibilities of these new technologies to present their learners with texts and learning task materials that are much more life-like, topical and attractive than the constraints of traditional printed materials would allow. Secondly, current educational reforms and pedagogical trends in Europe encourage more independent and autonomous learning. Most current textbooks do not lend themselves easily to this use, so teachers need to adapt or develop their own (self-access) materials for their learners. Furthermore, there is a growing demand on teachers to differentiate. Published materials are by their nature generally less suitable for this purpose. In such circumstances, as Marsden (2001) has shown in his studies of textbooks, teachers will often have to adapt or design supplementary course materials. Finally we see a trend towards competence-based learning. In Europe this development has received an impetus by the introduction of the Common European Framework of Reference (CEFR; Council of Europe 2001; Little 2007). However, as Westhoff (2007) argues, the CEFR is not easily compatible with traditional grammar-oriented pedagogy, as reflected in the most commonly used textbooks. Those

Revised version of a plenary lecture presented on 12 February 2008 in the series Assessing Language Learning at the Language Institute, University of Wisconsin–Madison.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

practitioners who want to shape their work according to principles underlying the CEFR will have to look for alternatives to the usual form-focussed textbook tasks.

These four developments demand from teachers an extended competence in evaluating learning materials. This is especially true of tasks that are life-like and content-oriented and, as a consequence, integrate many aspects of language acquisition. However, as de Matos (2000) points out, the critical analysis of such materials has often been overlooked or even neglected in teacher education programmes. In order to fill this gap, a scaffolding instrument for *a priori* assessment of the learning impact of language learning tasks seems to be useful. In this lecture I will describe the development of such an instrument, its theoretical basis, some of its characteristics and the subsequent stages we have gone through in our endeavour to develop a feasible, reliable and valid form for it, resulting in the creation of different tools for different purposes.

2. A 'yardstick' for complex language learning tasks

The incentive for the development of a tool for the assessment of language learning tasks was the fact that Dutch FL teachers looking for learning tasks that were realistic and embedded in rich, life-like contexts in which they were functional discovered WebQuests (Dodge 1995). Dodge defines the concept as follows: 'A WebQuest is an inquiry-oriented activity in which some or all of the information that learners interact with comes from resources on the internet' (http://webquest.org/index-resources.php). Dodge gives a set of well thought out, generic criteria to assess the quality of WebQuests. However, WebQuests are developed for all subjects in the current school curricula. Unfortunately, these generic criteria do not provide indications about those characteristics of tasks which specifically foster second language acquisition (SLA). Thus, the use of some WebQuests may contribute substantially to SLA, whereas others do not yield much progress. Our first aim was to help these teachers choose and/or adapt or design WebQuests that could be expected to foster SLA by developing an instrument for qualitative analysis. We created a new name for those WebQuests, which appeared to focus on SLA: LanguageQuests (LQ). However, during the development of this instrument we noticed that we actually defined criteria that could be applied to many types of integrated, life-like and functional tasks.

2.1 The theoretical framework of the first 'yardstick'

In our view, we considered a first requirement to be that the instrument should be theory-based. We started with current insights from SLA theory in general and, more specifically, with the characteristics of tasks in a task-based approach (TBA). In general, no-one in cognitive psychology seriously doubts these days that acquisition of knowledge is the product of mental activity (see Anderson 2004). Specifically, knowledge is assumed to be stored in long term memory (LTM) in the form of traces resulting from mental actions in working memory (WM) (for an overview, see Baddeley 1997, 2001, 2007; Carroll 2004). According to Baddeley (2001: 852), the working memory is the place where information, coming from the sensory stores

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

and from LTM, is combined and assembled in a process of constructing meaning. This means that learning is basically activity-based and that a decisive part of this activity takes place in WM. Accordingly, tasks can be analyzed with respect to the type, variety and intensity of the mental activity they elicit in WM.

2.1.1 Five domains of mental actions

We reviewed current SLA theory and distinguished five activity domains in which mental actions could be expected to contribute to SLA. In spite of fierce debates on various issues, a certain consensus seems to exist regarding five basic assumptions on domains of activity that can be supposed to contribute to SLA (for an overview see Lightbown & Spada 1999; Brown 2000; Richards & Rodgers 2001; Mitchell & Myles 2004).

1. Exposure to input. Without extended exposure to a rich input, there is little SLA (Krashen 1985). Although very few of Krashen's ideas could be confirmed empirically, and most of them have now been largely challenged (McLaughlin 1987; Ellis 1990; Norris & Ortega 2000; DeKeyser 2003), there seems to be a broad consensus in the recent scientific literature that being exposed to a rich FL input is a crucial, although not exclusive, prerequisite for SLA. In order to be functional, this input should not be too far above the actual competence level of the learner. Expressed in Vygotskian terms, this level could be described as the learners' 'zone of proximal development'. In the literature, this level is commonly referred to as 'i + 1', where 'i' stands for 'interlanguage'.

2. Content-oriented processing. Mere exposure is not enough. There seems to be little doubt that being exposed to input is only effective if the input is processed, or, in more practical terms, if the learner has tried to understand its meaning. Something meaningful has to be constructed in WM before it can be passed on to and stored in LTM (see e.g. Johnson 1982; Wesche 1993; Skehan 1998).

3. Form-oriented processing. There is far less agreement on the role of grammar, or so-called 'form–focused instruction' (for an overview, see Burgess & Etherington 2002; DeKeyser 2003, 2005; Doughty 2003). Although research outcomes have been presented recently which advocate the role of formal instruction (Housen & Pierrard 2005), support seems to be growing for the so called 'weak interface hypothesis' (Ellis 1990). This hypothesis tries to explain the paradox that extended content-oriented input processing, combined with form-focused instruction, leads to better results than input processing alone, but that taught grammar rules are seldom used in producing output. The weak interface hypothesis claims that part of the learners' output is rule-directed, but that we do not know the rules they actually use. Learners probably construct their own rules by formulating hypotheses about aspects of the form of the FL by processing input. This process of hypothesis formation is supposed to be stimulated by directing the learner's attention to form aspects of the input. Such instruction is characterized as 'Focus on Form', to be distinguished from explicit grammar instruction, which is labelled 'Focus on FormS' (Long 1991; Doughty & Williams 1998). Altogether, there seems to be growing agreement that some kind of form-awareness-raising activity can be supposed to foster SLA.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

4. (Pushed) output. Several researchers have argued for the facilitating and stimulating role of (pushed) output production ('pushed' in this context means that demands are being made for correct and appropriate use of the foreign language). According to the output hypothesis (Swain 1995; Swain & Lapkin 1995; De Bot 1996) it is assumed to enhance fluency, and it makes language learners conscious of their deficits, and so increases their motivation to learn. The output hypothesis claims further that pushed output contributes to form-orientation and gives the teacher or the communication partner the opportunity to provide corrective feedback (for an overview of its effect, see Spada 1997; Toth 2006). Experiments seem to confirm this claim (Nobuyoshi & Ellis 1993; Swain & Lapkin 1995). Two varieties of output can be distinguished. One part of our language utterances consists of unanalyzed combinations (chunks) that are perceived as wholes (Lyons 1968). Their use is labelled 'formulaic speech' (Myles, Hooper & Mitchell 1998). Pushed output increases the learners' ability to use these 'chunks' in different situations and combinations (Schmitt & Carter 2004; Schmitt et al. 2004). The other variety is somewhat misleadingly labelled 'creative speech' (Ellis 1986: 167–170). Misleading, because it has little to do with poems or creative writing; the term is used for rule-guided production.

5. Acting strategically. Generally speaking, only limited time is available for foreign language education. This means that there will always remain gaps in our knowledge. For that reason, it is useful to develop a repertoire of strategies to compensate for these deficiencies. We can compensate for deficiencies in receptive knowledge by applying reading and listening strategies such as inferring unknown elements, using prior knowledge, etc. (Westhoff 1991a, b). To make up for deficiencies in productive competence, we can use communication skills such as negotiating meaning, avoiding, description, fillers (Bialystok 1990; Poulisse 1990; Littlemore 2001).

These five assumptions can be considered as the basic ingredients of effective SLA activities. Together they form what Westhoff (2004) calls the 'SLA penta-pie', containing all the basic ingredients of a complete and nutritious 'language meal'.

2.1.2 Criteria from a task-based approach

There is an extended literature about tasks and the TBA. There are many different definitions of the concept of task, as well as differences in the claims regarding a task's effectiveness (for an overview, see Bygate, Skehan & Swain 2001; Ellis 2003). They vary from very broad definitions like Breen's description that in fact encompasses as good as all classroom activities directed at language learning 'form the simple and brief exercise type to more complex and lengthy activities such as group problem-solving or simulations and decision-making' (Breen 1987: 23) to Skehan's (1998) extensive definition which states, among other things, that tasks should be meaningful, related to real-world activities, and learners should not be asked to regurgitate other people's meanings. Willis (1996) emphasizes the need for goal-orientedness of a task, requiring that the target language has to be used for communicative purposes for an outcome to be achieved. According to Swain & Lapkin (2001) tasks should involve learners in interactive and collaborative activities in order to elicit negotiation of meaning.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

However, there is common core of characteristics for effective tasks; they are as follows:

- Tasks should be oriented to content. To understand and to be understood should be the main concern.
- Tasks should be life-like and functional in the sense that they produce an outcome in the form of a useful product.
- Learners should be allowed or even encouraged to make use of their own linguistic resources.
- The activity should be holistic.
- It should preferably be carried out as teamwork, mostly with divided roles.
- Although the product has to meet pre-set specifications, a considerable flexibility should be allowed in the way learners go about executing the process.

2.2 The construction of a 'yardstick'

In addition to the generic criteria for WebQuests as developed by their originator Bernie Dodge, we constructed two supplementary sets of criteria, one covering the five action domains presented earlier, inferred from SLA theory, and one where we aimed to summarize the task characteristics tapping into the extent to which the analyzed LanguageQuest could be considered to foster SLA according to the claims of the TBA.

The first version of the instrument was designed in the Netherlands in cooperation with a group of FL teachers whose main objective was the construction of new powerful WebQuests for SLA. The 'yardstick' was a by-product of the construction process. It had the form of a checklist consisting of statements. For outsiders the initial product appeared to be not very user-friendly and its use required more knowledge of SLA theory and terminology than practitioners usually possess. In the second stage we used a Delphi-like procedure. First we revised the instrument quite substantially on the basis of earlier experiences. This new tool was tried out by a group of teachers in a workshop setting. With the help of their (not very favorable) comments and suggestions for improvement, the wording of the criteria was substantially adapted again and the tool became an elaborate rubric. This third version was tried out anew by a further group of teachers for fine tuning. At this stage it became clear that, in order to be transparent, unambiguous and understandable for those who might not have read extensively in the field of SLA theory, the formulations of the descriptors had to be rather extensive. As a consequence the instrument became lengthy and too unwieldy for practical application. We tried to solve this problem by digitizing the 'yardstick' so that we could use a hypertext structure. As is demonstrated in Table 1, only the global descriptor of each criterion is visible at the surface. More detailed specifications and explanations are given in the underlying text, which can be reached on demand by clicking the surface text. In every extended description, significant keywords are highlighted (bold text in Table 1).

The final version was translated from Dutch into English and German. The English version was tried out in an international workshop of the European Centre for Modern

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

Table 1 Two examples of criteria for assessing LanguageQuests.

Criterion 12: Process with focus on form
Surface text: To what extent does the task stimulate (the further development of) the learner's language awareness (notice form aspects in the materials and possible regularities in them such as word order, tenses, word endings, prefixes and suffixes, plural forms, etc.).
Hypertext:

1	2	3	4
In **no** way does the task promote the learner's **attention** for aspects of form in the materials offered, such as word order, use of tenses, plural forms, etc.	The task does promote the learner's **attention** for aspects of form in the resources offered but is not functional: The task can be done without actual further exploration. It is a grammatical **exercise in its own right**, more or less by way of intermezzo.	The task requires awareness of the formal aspects of the language involved and **although this is also somehow related to the content aspects of the task the primary focus is on the grammatical/ syntactical phenomenon in itself**.	The task requires awareness of the formal aspects of the language involved but only to the extent that this **awareness contributes to a better performance in task completion**. The focus is on the impact of form on meaning. But learners are stimulated to pay close attention to language and form.

Criterion 18: Closed or open assignment?
Surface text: No further surface text
Hypertext:

1	2	3	4
The task and/or parts of the task ask(s) for one concrete result (Type: 'In which of the following restaurants can you get a hamburger?') There is **only one correct solution**, which was pre-determined. **Resources are supplied and mandatory**.	There is **some variation in outcome or product**, but the requirements for product and presentation are fairly rigid. There is **little variation in the final products to be expected, nor is it wished for**. Other resources than the mandatory ones may be used, **but the latter should at least be consulted**, often with set assignments.	**Many different results are possible**, but there is so much control that the **various products do not differ very much. The use of the supplied resources is recommended**. The use of these (or additional) resources is free.	Requests a product or parts of the product using specifications (Type: 'Agree on a menu. You can spend €10 each.') Essentially there is an **infinite number of different 'correct solutions'**. The way the students approach the problem is free. **The supplied resources are optional**. Originality is stimulated.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

Languages (ECML) in Graz (Austria) with satisfactory results (http://www.ecml.at/mtp2/Lquest/html/LQuest_E_Results.htm). The German version was used in a regular seminar for undergraduate students of L2 German at the University of Vienna. These students used the 'yardstick' both as an evaluation tool for existing WebQuests and as a structuring tool in the process of constructing new complex tasks. It was used by the students without support. The students were positive in their evaluation of both functions.

The final versions (in Dutch, English and German) were complemented with a grading system (one, two or three 'stars'). Subsequently, the 'yardstick' was put on the Internet in an interactive format (http://webquest.kennisnet.nl/talenquest/talenquest_beoordelen). Once the questions are completed, the end score per category and the total score, including the grading in terms of 'stars', are calculated automatically. Comments and suggestions for improvement of the reviewed task can be added by the reviewer.

3. An assessment tool as learning material for teachers

Although the LQ 'yardstick' appears to help practitioners in estimating an LQ's appropriateness for FL learning, it tells little about the nature of elicited mental activities and their relative effectiveness in terms of learning impact. Evidently, not all activities in the domain of 'form-oriented processing' found in current textbooks can be assumed to be equally effective for SLA. Although theory-based, our 'yardstick' did not provide an in-depth analysis, necessary to tap into those differences in effectiveness. An analysis that aims at determining the characteristics of mental actions elicited by a task would be so time-consuming that it seemed unfeasible to construct a tool that would cover an extensive and complex task like an LQ, let alone an extensive body of materials such as a complete textbook. On the other hand, experiences with early versions of the tool led to the supposition that working with an instrument for in-depth analysis of relatively small tasks enhanced one's competence to make 'educated guesses' about the potential effectiveness of more complex ones. Users seemed to develop a keen eye for the learning impact to be expected from tasks, so we decided to explore the possibilities of a tool for in-depth analysis of small tasks which would serve as an instrument of learning for teachers, aiming at the enhancement of their competence in estimating the impact of language learning tasks. For this purpose, we had to develop a new tool because existing ones did not provide assistance in delineating the characteristics of mental actions in WM and their role in constructing and storing meaning. Such tools do not contribute to enhancing a teacher's insight into which characteristics of a task promote learning. And this insight was, in our view, precisely the basis of their increased competence. We needed a theoretical framework that would give us more clues regarding those characteristics of mental actions that account for learning results.

In tackling this problem we used the Multi-Feature Hypothesis, as presented by Westhoff (2004) and further developed and refined by Moonen, De Graaff, & Westhoff (2006). In this hypothesis, insights are combined from schema theory (Rumelhart & Ortony 1977) and connectionism (McClelland, Rumelhart & Hinton 1986; Gasser 1990; Bereiter 1991, 2002; Plunkett 1998). Among cognitive psychologists, it is widely accepted that the product of a cognitive learning process should not be perceived as a template, but as a more or less open

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

mental structure of neural units (Greeno & Simon 1993). Some call it a 'structure of features' (Klausmeier & Allen 1978), others a 'schema' (Rumelhart & Ortony 1977; Rumelhart 1980), and others still 'associative networks' (Bereiter 1991). Anderson (2004: 22) writes: 'We can be sure that human cognition is achieved through large patterns of neural activity'. Such patterns or networks are not necessarily distinct entities. According to Gasser (1990), the network structure of a concept is distributed over many units, each of which can also participate in the representation of many other concepts. Properties or characteristics of concepts consist of combinations of such units. For such combinations the term 'features' is often used. Features can be linguistic as well as non-linguistic. The identity of a concept consists of a distinctive set of features.

We find parallel views among cognitive psychologists concerning the way knowledge of such patterns of combinations is formed. Gasser (1990) and Anderson (2004), for example, suggest that they emerge by having been (repeatedly) processed in combination with each other in WM. According to this connectionist theory, our brain keeps track of the regularities in the occurrence of combinations and of the frequency of occurrence of these combinations in WM. The frequency determines the 'weight' of the established connections between the features. This 'weight' accounts for the ease of retrieval from LTM. On the basis of these insights, the Multi-Feature Hypothesis (MFH) predicts that retention and ease of activation are improved by mental activities involving many features from many different categories, manipulated in such a way that they occur in current combinations, in great frequency, and simultaneously. This led us to the design of a typology of tasks, distinguishing task types in such a way that it would provide information about the estimated amount, variety and frequency of features (linguistic and non-linguistic) apparently involved in the elicited mental action. In 'ranking according to a given criterion' (such as in ordering based on chronology), for example, only one feature has to be considered (point in time), whereas in 'generalizing' (such as deducing a rule from a series of cases), many features of different kinds have to be considered (in order to establish which ones have enough in common to be considered as a group under a single heading). Combinations of these actions may be possible in carrying out a specific task. Based on these considerations, seven mental actions were defined (see Table 2).

We then developed a scaffolding instrument containing a set of theoretically-motivated questions and guidelines by which the users were guided to determine systematically the activity domain covered by a task and the type of activity elicited by the assignment (for a full description, see Driessen et al. 2008). Our assumption was that having to choose the task type would elicit the mental action of examining the assignment to be analyzed regarding variety of features to be processed. This in turn would yield awareness of, and deeper insight into, the role of feature variety in a mental learning action. We evaluated the instrument in the form of a systematically monitored trial using both quantitative and qualitative methods. The results show that, even after brief training, teachers reported that the use of the tool contributed to their competencies. Most teachers were satisfied with the new knowledge base they said they received. The data show that the teachers were not only able to work with the tool, but also reported that they gained knowledge and insight into which characteristics of tasks account for learning effects (Driessen 2003). However, it should be borne in mind that the outcomes of the evaluation also show the need for a quite extensive clarification of

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

Table 2 *Mental actions in SLA with descriptions, explanations and illustrations (from Driessen et al. 2008).*

Mental action	Description	Explanation and illustration
Reproducing	Repetition of the language input	In language education it is a common procedure to elaborate language input on the basis of repetition, rehearsing and memorizing such as remembering words by listening and repeating aloud.
Ranking	Positioning the language input by virtue of a shared feature that is ordinal or taxonomic in nature	Possible ways of ranking: • simple, such as based on chronology or meaning • complex, such as based on use or attractiveness
Categorizing	Arranging the language input into groups	Examples of categorizing: • What belongs to content, relation, procedure? • What do you recognize in . . .? Do you agree with . . .? For what groups could it be interesting to . . .? • Search (according to some description) . . . • looking for similarities and differences
Structuring	Putting the language input into one or more groups and determining their relationships	Examples of structuring: • grouping words according to their meaning by making 'word maps'
Generalizing	Taking the language input to a level more general than the one given	Examples of generalizing: • deduce a rule from some concrete instances • recognize a pattern • note taking; summarizing • devise questions with texts
Applying	Taking the language input to a more concrete level than the one given	Examples: • think of examples or instances of a general principle or idea • think of illustrations • think of directions for use
Elaborating	Relating the language input to knowledge already stored in the memory	For example, answering questions such as: • What do I know already? • What else do I need? • Where and how can I find it? • What's its relevance?

the concepts. This must include theoretical background information in concrete terms and practical illustrations. Furthermore, successful use requires some introductory training and guidance.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

4. The search for reliability and validity

4.1 Variety and frequency of features

The tool we had developed appeared to be useful as a learning task for practitioners. However, although it provided insight about variety and frequency of features that are probably or potentially involved in the elicited mental action, these insights remained rather general and non-specific. Moreover, the fact that teachers who had completed a training in the use of this tool reported enhanced competence in estimating the learning impact of a task does not prove that they were right in their judgements. Even the fact that they agree in their judgements is no evidence for the actual effectiveness of the assessed tasks. Accordingly, we decided to investigate the possibility of designing an instrument that enables the user to predict the quality of a task in a more reliable and valid way. Again we took the MFH as the theoretical basis but used it in a more detailed way. First, according to connectionist theory a learning result consists of a network of features. Features in a neural network activate each other and activation of a network can start from any connected feature. From that it can be concluded that networks consisting of many different features from different (linguistic and non-linguistic) categories have more chance of becoming activated in diverse situations by manifold evocators. Hence, a variety of features seems to be one important characteristic of a learning activity. Secondly, retention and ease of activation depends on the 'weight' of a connection. 'Heavy' connections yield prominent features. The 'weight' of connections is determined by the frequency of their presence in WM. Combinations will be easier to retrieve from LTM the more frequently they are processed in WM; therefore, the frequency of being processed in WM is another important factor. Hence, we decided to get a closer look at the variety and frequency of features processed in WM. We tried to get a better grasp of the variety of features involved in a mental action by describing a number of linguistic and conceptual categories, and distinguished the following LINGUISTIC feature categories (for a full description, see Moonen 2008):

- ORTHOGRAPHICAL FEATURES, carrying information about different aspects of the written word form.
- PHONOLOGICAL FEATURES, containing information about the phonological structure.
- SYNTACTIC FEATURES, dealing with 'what part of speech it is and what grammatical patterns it can fit into' (Nation 2001:55).
- MORPHOLOGICAL FEATURES, giving information about various alternative forms of words.
- SEMANTIC FEATURES, containing information about the linking of the correct labels of the L1 and the L2, in either direction.
- COLLOCATIONAL FEATURES, containing information about what words frequently co-occur with the target item, at the sentence level, including unanalyzed chunks.

We also described a number of CONCEPTUAL feature categories that can participate in the representation of a target item and that belong to general world knowledge or our personal knowledge of specific events.

- PRAGMATIC FEATURES, referring to what a concept can be used for or what can be achieved with it.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

- SENSORY FEATURES, relating to a wide range of sensory aspects of a target item, such as visual appearance, color, taste, texture and sound.
- AFFECTIVE FEATURES, referring to personal emotional or evaluative associations, e.g., positive and negative emotions, feelings and personal experiences associated with a target item.
- SITUATIONAL FEATURES, containing information about the general, non-linguistic context or environment in which a concept commonly or frequently occurs.

In order to be able to estimate whether such features from these categories are processed, we formulated a working definition of each feature category. The working definitions were tried out a few times until we had formulations that appeared to be satisfactorily distinctive. Processing morphological features, for example, was assumed to take place when inflections or derivations of a word had to be processed. For the *a priori* estimation whether processing of feature categories would probably be elicited by the task, we designed a task sheet in the form of a matrix in which for each task element the feature categories which will possibly be involved in the elicited mental action can be indicated (Moonen 2008). In this way the tool is also usable for (relatively small) composed tasks.

4.2 Design and procedure

In order to test the reliability and the validity of the instrument we designed two versions of a vocabulary learning task: a 'rich version', in which the learners were assumed to process a broad variety of features, and a 'poor' one, which was assumed to elicit only a limited number of features. In the former, the learners were asked individually to categorize a given set of German words into groups that can form the outline of a micro-story and subsequently, in pairs, to choose the best outline for such a story, work out an appropriate formulation for it, write it down in German and memorize it by rehearsing the story in dyads. In the 'poor' version the students were given the German words with their Dutch translations and asked to memorize them individually and to rehearse them afterwards in pairs. Both tasks were judged *a priori* by two raters with the help of the task sheets. Both percent agreement and interrater reliability (Cohen's kappa) were calculated. Using the labels attributed to the relative strength of agreement proposed by Landis & Koch (1977: 165), the kappa statistics of the orthographical, phonological, syntactic, morphological, collocational and situational feature categories were considered 'almost perfect'. The kappa of the semantic and pragmatic features were considered 'substantial'. The kappa of the sensory features were labelled 'fair' and the kappa of the affective features 'slight' (Moonen 2008). We believe that these results sufficiently confirm the instrument's reliability for practical use.

However, these results do not prove the superior learning impact of the task that was estimated to be 'rich'. For this purpose we measured retention of the words presented in the analyzed tasks. For that purpose both tasks were carried out in intact classes of a Dutch secondary school (for the 'rich' task $N = 46$, for the 'poor' task $N = 44$). One week before the experiment took place, a pre-test was administered to measure prior knowledge of the vocabulary to be presented. Learning gains were measured by a translation test and a written

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

description test in which the students were asked to describe situations in a cartoon. The translation test examined whether the learners were able to reproduce the target words in a situation which provided minimal contextual support and which was expected to make minimal demands on WM. The written description test was used to examine whether the learners were able to use the target words in a relatively rich context, which was expected to make considerable demands on WM capacity. Two days after the experiment, a direct post-test was administered, one week thereafter a delayed post-test. We administered an ANCOVA with the pre-test as covariate.

4.3 Outcomes

Results show that the 'rich'-task group significantly outperformed the 'poor'-task group ($p < .05$) on both test types and on both direct and delayed tests (Moonen 2008). On the written description test the differences appeared to be greater than on the translation test. This can be interpreted as further indication that the tasks that were rated 'rich' indeed yielded learning results that can be more easily activated in complex, life-like situations.

5. Some conclusions

Our research shows that it is possible and profitable to delineate a systematic way of assessing the impact of language learning tasks on a sound theoretical footing. Indeed, practitioners who are not experts in applied linguistics appear to be able to apply a tool based on principles derived from SLA theory and cognitive psychology. Furthermore, these results confirm our expectation that our instrument for assessing learning tasks *a priori* predicts the potential learning impact of such tasks in a reliable and valid way. However, such a reliable and valid in-depth analysis is cumbersome and time-consuming. In order to prove reliability and validity of the instrument as described in section 4, we had to restrict ourselves to relatively simple tasks (the acquisition of 16 words). It remains questionable whether such an instrument is very workable in everyday teaching practice. The 'yardstick', as described in section 2, may offer the best solution for practical use. Although in only scratches the surface of what happens in WM, it has a sound theoretical basis and allows for the assessment of differentiation in learning tasks. As was shown in our second experiment, this profit can be enhanced by incidentally performing a more in-depth analysis of smaller tasks with the help of the instruments described above, because it can be expected that this will contribute to the heightening of awareness in these matters. As a consequence it can be expected that the incidental performing of in-depth analyses would make the use of the 'yardstick' more trustworthy.

Assessing the effectiveness of language learning tasks appeared to require a certain, albeit elementary, knowledge about SLA and cognitive psychology. In our second experiment, in spite of the extensive and explicit wording of the criteria, its use required an introductory training session with a detailed clarification of the concepts used. Even in the restricted and very concrete and operationalized form of the instructions and criteria of our last experiment,

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

one of the raters, a trainee teacher of L2 English needed a short supplementary training session before he could carry out the analyses. Given the fact that the significance of teachers' competence in estimating the learning impact of instructional materials is increasing, this finding underscores the importance of SLA theory and cognitive psychology instruction in language teacher education.

References

Anderson, J. R. (2004). *Cognitive psychology and its implications* (6th edn.). New York: Worth.
Baddeley, A. D. (1997). *Human memory: Theory and practice* (revised edn.). Hove: Psychology Press.
Baddeley, A. D. (2001). Is working memory still working? *American Psychologist* 56, 849–864.
Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
Bereiter, C. (1991). Implications of connectionism for thinking about rules. *Educational Researcher* 20, 10–16.
Bereiter, C. (2002). Liberal education in a knowledge society. In B. Smith (ed.), *Liberal education in a knowledge society*. Peru, IL: Carus, 11–34.
Bialystok, E. (1990). *Communication strategies*. Oxford: Blackwell.
Breen, M. P. (1987). Language learning tasks. In C. N. Candlin & D. F. Murphy (eds.), *Learner contributions to task design*. London: Prentice Hall, 5–23.
Brown, H. D. (2000). *Principles of language learning and teaching* (4th edn.). White Plains, NY: Pearson.
Burgess, J. & S. Etherington (2002). Focus on grammatical form: Explicit or implicit. *System* 3.4, 433–458.
Bygate, M., P. Skehan & M. Swain (eds.) (2001). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Pearson.
Carroll, D. W. (2004). *Psychology of language* (4th edn.). London: Wadsworth.
Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning* 46.3, 529–555.
DeKeyser, R. M. (2003). Implicit and explicit learning. In Doughty & Long (eds.), 313–348.
DeKeyser, R. M. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning* 55.1, 1–25.
Dodge, B. (1995). Some thoughts about WebQuests. http://WebQuest.sdsu.edu/materials.htm.
Doughty, C. J. (2003). Instructed SLA: Constraints, compensation and enhancement. In Doughty & Long (eds.), 256–310.
Doughty C. J. & M. H. Long (eds.) (2003). *Handbook of second language acquisition*. Oxford: Blackwell.
Doughty C. J. & J. Williams (eds.) (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.
Driessen, C. (2003). *Analyzing textbook tasks and the professional development of foreign language teachers*. Utrecht: WCC.
Driessen, C., G. Westhoff, J. Haenen & M. Brekelmans (2008). A qualitative analysis of language learning tasks: The design of a tool. *Journal of Curriculum Studies* 40.1, 1–18.
Ellis, R. (1986). *Understanding second language acquisition*. Oxford: Oxford University Press.
Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Blackwell.
Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition* 12.2, 179–199.
Greeno, J. G. & H. A. Simon (1993). Situativity and symbols: Response to Vera and Simon. *Cognitive Science* 17, 4.
Housen, A. & M. Pierrard (eds.) (2005). *Investigations in instructed second language acquisition*. Berlin & New York: Mouton de Gruyter.
Johnson, K. (1982). *Communicative syllabus design and methodology*. Oxford: Pergamon.
Klausmeier, H. J. & P. S. Allen (1978). *Cognitive development of children and youth: A longitudinal study*. New York: Academic Press.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.
Landis, J. R. & G. G. Koch (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
Lightbown, P. M. & N. Spada (1999). *How languages are learned* (revised edn.). Oxford: Oxford University Press.
Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language educational policy. *The Modern Language Journal* 91.4, 645–655.
Littlemore, J. (2001). An empirical study of the relationship between cognitive style and the use of communication strategy. *Applied Linguistics* 22.2, 241–265.
Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. D. Bot, R. Ginsberg & C. Kramsch (eds.), *Foreign language research in cross-cultural perspective*. Philadelphia, PA: John Benjamins, 40–52.
Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
Marsden, W. E. (2001). *The school textbook: Geography, history and social studies*. London: Woburn Press.
Matos, G. de (2000). Teachers as textbook evaluators: An interdisciplinary checklist. *IATEFL Issues* 157, 7–9.
McClelland, J. L., D. E. Rumelhart & G. E. Hinton (1986). The appeal of parallel distributed processing. In J. L. McClelland & D. E. Rumelhart (eds.), *Parallel distributed processing: Explorations in the ministructure of cognition* (vol. 1). Cambridge, MA: MIT Press/Bradford Books, 3–44.
McLaughlin, B. (1987). *Theories of second-language learning*. London: Edward Arnold.
Mitchell, R. & F. Myles (2004). *Second language learning theories* (2nd edn.). London & NewYork: Arnold.
Moonen, M. (2008). *Testing the multi-feature hypothesis: Tasks mental action and second language acquisition*. Utrecht: IVLOS Series.
Moonen, M., R. De Graaff & G. J. Westhoff (2006). Focused tasks, mental actions and second language teaching: Cognitive and connectionist accounts of task effectiveness. *ITL – International Journal on Applied Linguistics* (special issue: *Task-based Language Teaching*) 152, 35–55.
Myles, F., J. Hooper & R. Mitchell (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48.3, 323–362.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nobuyoshi, J. & R. Ellis (1993). Focussed communication tasks and second language acquisition. *ELT Journal* 47, 203–210.
Norris, J. & L. Ortega. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50.3, 417–528.
Plunkett, K. (1998). Language acquisition and connectionism. In K. Plunkett (ed.), *Language acquisition and connectionism*. Hove: Psychology Press, 97–105.
Poulisse, N. (1990). *The use of compensatory strategies by Dutch learners of English*. Dordrecht: Foris.
Richards, J. C. & T. S. Rodgers (2001). *Approaches and methods in language teaching* (2nd edn.). Cambridge: Cambridge University Press.
Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. S. Brewer (eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education*. Hillsdale, NJ: Erlbaum, 33–58.
Rumelhart, D. E. & A. Ortony (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro & W. E. Montague (eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum, 99–135.
Schmitt, N. & R. Carter (2004). Formulaic sequences in action: An introduction. In N. Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 1–22.
Schmitt, N., Z. Dörnyei, S. Adolphs & V. Durow (2004). Knowledge and acquisition of formulaic sequence: A longitudinal study. In N. Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 55–86.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching* 30, 73–87.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (eds.), *Principles and practice in applied linguistics: Studies in honour of H. G. Widdowson*. Oxford: Oxford University Press, 125–144.
Swain, M. & S. Lapkin (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics* 16, 371–391.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press

Swain, M. & S. Lapkin (2001). Focus on form through collaborative dialogue: Exploring task effects. In Bygate et al. (eds.), 99–118.
Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning* 56.2, 319–385.
Wesche, M. (1993). Discipline-based approaches to language study: Research issues and outcomes. In M. Krueger & F. Ryan (eds.), *Language and content*. Lexington, MA.: D. C. Heath, 80–95.
Westhoff, G. J. (1991a). Increasing the effectiveness of foreign language reading instruction (part 1). *ADFL Bulletin* 22.2, 29–36.
Westhoff, G. J. (1991b). Increasing the effectiveness of foreign language reading instruction (part 2). *ADFL Bulletin* 22.3, 28–32.
Westhoff, G. J. (2004). The art of playing a pinball machine: Characteristics of effective SLA tasks. *Babylonia* 12.3, 58–63.
Westhoff, G. J. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal* 91.4, 676–679.
Willis, J. (1996). *A framework for task-based learning*. Baisingstoke: Longman.

GERARD WESTHOFF is Professor of L2 pedagogy at Utrecht University (IVLOS Institute of Education) in the Netherlands. As a teacher educator he currently works mainly in research and in-service training. His research focuses on instructional design, more specifically on task design for L2 acquisition. He is director of the National Bureau on Modern Languages, created and financed by the Dutch Ministry of Education. The Bureau's mission is to improve the quality of modern language learning and teaching in the Netherlands.

https://doi.org/10.1017/S0261444808005697 Published online by Cambridge University Press