In: P. Seedhouse (ed) (2017) *Task-based Language Learning in a real-world Digital Environment: The European Digital Kitchen*. London: Bloomsbury, pp. 207-229.

**Chapter 10**

# Vocabulary Learning in a Real-World Digital Environment

*Gabriele Pallotti and Natacha Niemants, University of Modena and Reggio Emilia, Italy*
*Paul Seedhouse. Newcastle University, UK*

***Chapter Summary***

*This chapter reports on how learning of vocabulary items was assessed in the English and Italian digital kitchens, the methods being different but complementary. In England a quantitative approach was adopted, involving pre-test, post-test and delayed post-test of a set of 16 vocabulary items (utensils and ingredients) learnt by individuals. The Italian team followed a broadly similar research design, with two main differences: there was no delayed post-test, and the vocabulary items presented in the pre- and post-test were different, to control for possible test-retest effects. The results of both studies showed gains resulting from the learning sessions in the European Digital Kitchen (EDK), which reached statistical significance in the English data. The combination of the two studies provides a rich and detailed perspective on one single aspect of the overall learning experience in the EDK.*

## Assessing learning in an immersive digital environment

One problem for any holistic environment for language learning is how to assess language learning precisely. A pervasive digital language learning environment is intended to be a holistic one, in which learners autonomously access resources to complete a task and thereby learn aspects of a language as well as other skills. However, this does pose certain problems when it comes to the precise evaluation of the effectiveness of such an environment. Exactly which aspects of a language have been learnt? How do we know participants did not know an item previously and what is the evidence that it has now actually been learnt? More generally, if we are trying to create an autonomous, holistic environment, would this not be disrupted by testing procedures? Ideally, the evaluation of a holistic environment would itself be holistic and evaluate all aspects of language learning together. In chapter 3 we provided a holistic illustration of learning processes in the French Digital Kitchen by presenting representative episodes from a complete task cycle.

In this chapter, by contrast, we decided on a narrow focus on one specific component of language learning for evaluation. This would enable us to see whether there was concrete evidence of learning in one narrowly delineated component of the overall language learning system. The main research question was: to what extent does learners' ability to verbally produce specific vocabulary items change as a result of a cooking session in this pervasive digital environment? The basic research design (described below) was a pre-test/post-test of specific vocabulary items, carried out on 50 learners of English and 34 learners of Italian. The intervention which was intended to promote learning of the items was the complete experience of a cooking session in the European Digital Kitchen, lasting about an hour.

Ellis (2012: 230) lists three ways in which tasks can be evaluated: 1) student-based evaluation, which consists in asking students' opinions about the task they have been engaged in; 2) response-based evaluation, based on the researcher's observation of how the task developed and whether it elicited the intended performance and outcomes; 3) learning-based evaluation, assessing students' abilities before and after the task to check whether any learning has occurred. All three forms of evaluation were included in a typical cooking session of the European Digital Kitchen (EDK): students were debriefed after the task, in order to gather their reactions; their performance on tasks was recorded and analysed; their lexical skills were tested before and after the session. This chapter reports mainly on the third dimension, learning-based evaluation, and, secondarily, on the second, response-based evaluation, which is more thoroughly analysed in other chapters in this volume. Data on the first dimension, student-based evaluation, have been collected, but are not presented in this collection.

**Learning vocabulary through actions and material objects**

One of the fundamental issues of L2 vocabulary research is defining the nature of learners' lexical knowledge and identifying valid and reliable ways of measuring it. Defining vocabulary knowledge in terms of 'the number of words a person knows' looks intuitively reasonable, but raises two important methodological questions: 1) how can one practically assess the whole set of lexemes known by an individual, which may easily reach several thousands?; 2) what does it mean to know a word?

As regards the first issue, research on vocabulary size usually consists in probing the receptive knowledge of small sets of lexical items taken from different frequency bands (e.g. the first 1,000, 2,000, 3,000 etc. words, ranked according to their frequency in a corpus), or looking at their presence in spontaneous productions, and using the results of these sample-based observations to draw generalising inferences about an individual's whole vocabulary. This type of analysis does not concern us here, however, for our aim will not be establishing the size

of our participants' global vocabularies, but whether they show any change in their knowledge of a small set of lexical items related to cooking.

The second methodological question is more relevant to our endeavour. In fact, in order to be able to state whether participants 'know more words' after taking part in the cooking sessions, one has to be clear about what is meant by 'knowing a word'. To answer this question, Nation (2001) proposed an influential framework based on three main dimensions: form (how the word is written and pronounced, and how it can be decomposed into smaller parts); meaning (how the word expresses concepts and how these are related to others in semantic networks); use (how the word can be used in association with others, its collocations, co-occurrence constraints and grammatical functions). Each of these dimensions can in turn be investigated in reception and production. However, according to some scholars, the receptive/productive dichotomy itself is too simplistic and they have proposed more complex frameworks to better articulate these notions. Schmitt (2010), elaborating on a proposal by Laufer and Goldstein (2004), suggests that word knowledge can concern meaning and form, and that it can be tested in terms of recall and recognition. Hence, given a word's meaning (e.g. presenting an object or a picture, or an L1 term), producing its form in the L2 would amount to recalling it, while selecting the correct form among a small set of alternatives (e.g. 4-5 words) would be an instance of form recognition. On the other hand, if a learner is presented with the form of a given L2 word, meaning recall would consist in being able to explicate its meaning, e.g. by drawing the object referent or providing an L1 translation, while meaning recognition could be demonstrated e.g. by selecting, among a small set of pictures/objects or words/phrases in the L1 or L2, the one corresponding to the word's meaning. Following this terminology, the tests used in our project can be deemed to be about form recall. This, according to Laufer and Goldstein (2004), is the highest level of word knowledge, implying the ability to recall from long-term memory the L2 form corresponding to a given meaning.

Most vocabulary tests are administered on paper or on a computer screen, and word meaning is thus operationalised with pictures or definitions in the L1 or the L2. Our tests of vocabulary knowledge are different, in that learners are given words' meanings by being presented with concrete material objects, of which they have to recall the L2 form. This design is coherent with the global approach of the EDK and makes the testing phases more integrated with the whole task. The cookery task, in fact, delivers a vivid kinesic, multimodal experience, which enables powerful, integrated verbal and nonverbal memory links and involves all of the senses. Long (2015: 306) suggests that the combination of mental and manual work in tasks creates the conditions required for incidental learning.

This resonates with the notion of 'mediation', which is central to Vygotsky's theory of mind (Vygotsky [1934] 1986; Lantolf 2000). According to Vygotsky, human higher cognitive functions, such as language and abstract reasoning, emerge from a number of experiences mediated by concrete objects. This holds true phylogenetically (in the evolution of our species), ontogenetically (in child development), and microgenetically (the many instances of learning something new in our life). The material object is thus an external aid in the way

to internalize thinking, allowing individual, higher-order cognitive processes to take place through social and semiotic mediation. For sure, adult learners can learn new words in ways that do not imply any manipulation of concrete objects, as for example through definitions or illustrations, and there are many words expressing abstract meanings that simply have no concrete referents. However, it has been claimed that even abstract words are rooted in bodily image-schemas (Jirak *et al.* 2010) and that all human cognition, including language, is embodied (Lakoff and Johnson 1999).

It would thus seem plausible that words referring to concrete objects or actions were more easily acquired than more abstract ones. However, little research has been done to confirm this assumption. Laufer (1997: 148) and Allen and Vallette (1972: 114) discuss the intrinsic properties of a word which affect its learnability. N. Ellis and Beaton (1993: 565-6) report that nouns are the easiest word class to learn, that concrete nouns are learnt more easily than abstract ones and that items are learnt more quickly and effectively if nonverbal referents are used as stimuli. Similar results were obtained by de Groot (2006), who found that concrete words are learned much better than abstract ones. Furthermore, although L1 and L2 words seldom map exactly onto each other (Schmitt and McCarthy 1997: 2), with such concrete objects as 'milk', it is relatively straightforward to comprehend referential meanings. De Groot (2006) also notes that L2 translations of concrete L1 words are easier to acquire than abstract ones.

There is however some uncertainty in the literature as to the definitions and boundaries of 'concrete' and 'abstract' words. To avoid uncertainty, a 'concrete word' is defined in the current study as a word which refers to a physical object which can be physically touched and manipulated by the learner during a language task or test. Relatively little attention has been paid in the field of vocabulary learning to the question of whether it is easier to learn the names of concrete physical objects when they are physically manipulated by learners as part of a task, rather than simply observed visually at a distance. However, Nattinger (1988: 67) reports on Total Physical Response and suggests that these '… associations of vocabulary with physical actions have a dramatic effect on memory because students must commit themselves to the learning task by performing appropriate actions.' Oxford and Scarcella (1994: 240) also suggest that such activities are particularly appropriate to kinesthetic learners. So overall the vocabulary items were relatively easy to learn, but the requirement of active oral production of items was relatively difficult. Melka (1997: 98) explains the problems involved in getting a subject to produce a target word to check productive knowledge; presenting the subject with an adequate context is the major problem. In this project, however, the context was supplied by presenting the concrete objects in an authentic kitchen setting, and this may have mitigated the difficulty.

**Vocabulary Input in the EDK Task Cycle**

We now consider how the task cycle within the EDK is intended to provide input to vocabulary learning for the learners. The task cycle is separate from the test cycle, although one is wrapped around the other (Figure 1).

**[INSERT FIGURE 1 NEAR HERE]**

As we saw in chapter 3, the task structure consisted of pre-task, main task and post-task. In the **pre-task**, the system introduces the learners to vocabulary items needed in the main task by instructing them verbally to collect the corresponding object from a different area of the kitchen. If the learners do not understand the word spoken by the system, they may ask for help by moving the green interaction tool (Figure 2). This provides a verbal repetition (spoken more slowly) on first request and then a verbal repetition with a photograph with a written caption on second request (Figure 3). This ensures receptive recognition of each vocabulary item. Learners therefore have the opportunity to use both the 'guessing from context' and the 'explicit teaching' methods of vocabulary learning (Schmitt and McCarthy 1997: 3).

**[INSERT FIGURES 2,3,4 NEAR HERE]**

Following its introduction in the pre-task, each vocabulary item is then repeated verbally by the system at least once during the **main task** (the cooking session) as part of the cooking instructions, thus providing further input. At each point of the cooking session, learners may also request help, which would come in three steps: the first two are like those detailed above (a slow repetition of the initial prompt; if needed, a slow repetition accompanied by a picture and a written subtitle), while the third consists in a video clip showing the action to be performed (Figure 4). The participants may also produce the vocabulary items when speaking to each another as they conduct the task, as in Extract 1 below. The system therefore provides a basis for the learners to both recognise and produce the linguistic form which relates to a specific object. The system requires the learners to physically manipulate the objects during the tasks, whilst the task design provides the opportunity (but not the necessity) for participants to employ the vocabulary in their joint dialogue. In the **post-task**, the participants sample and evaluate the food that they have cooked. This gives them a further opportunity (but not obligation) to employ vocabulary learnt. So each learner hears the name of each vocabulary item a minimum of 2 times from the system, but there is no maximum. Learners can continue asking the system to repeat the name of an object as many times as they choose, and this particular word may occur an indefinite number of times in their oral interactions.

**Vocabulary Learning Processes in the EDK**

How do the test and task cycles (Figure 1) relate to what is known about vocabulary learning processes? The **pre-test** establishes in the participants' minds a need to learn the vocabulary items. If they do not know the item, they should realise that they will need to learn the item in the pre-task, and so an information gap and a motivation is created. In the **pre-task**, participants are required to match the word to the object by collecting the object, using the help facility where necessary. This provides the opportunity for them to learn words that they do not know. Attention is on the spoken form, with the goal of getting learners to be able to recognise a word when they hear it, and to be able to pronounce a word correctly (Nation 2001: 98). However, seeking help from the system will also supply them with the written form of the item. During the **main task** (cooking session), they are required to locate and manipulate the objects, matching word to object for a second time, thus reinforcing the learning of items. They also have the opportunity to use the words when communicating with each other on-task if they wish, whilst manipulating the objects. The **post-task** gives them a further opportunity (but not obligation) to employ vocabulary learnt whilst tasting and evaluating the dish. During the immediate **post-test**, participants are able to evaluate which vocabulary items they have learnt, following the identical assessment procedures to the pre-test, whilst the **delayed post-test**, when present, allows them to evaluate which items they have retained in the long term.

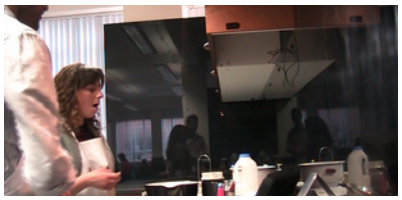
Nation (2001: 63) suggests that three processes lead to remembering a word, namely noticing, retrieval and creative use. The task and test cycles as a whole require learners to '…notice the word, and be aware of it as a useful language item…' (Nation 2001: 63). The post-test and delayed post-test require learners to retrieve the word multiple times. The learners have the opportunity (but not the obligation) to use the item creatively whilst performing the main cooking task, if they wish. Laufer and Hulstijn (2001) introduced the construct of task-induced involvement, which has the dimensions of Need, Search and Evaluation. They suggest (2001: 14) that *Need* is strong when imposed on the learner by him- or herself. The system does not tell learners that they need to learn vocabulary items. Rather, the pre-test should make them aware of any items they will need to employ, but which they do not know the names of, thus prompting the awareness of a need. Laufer and Hulstijn (ibid.) further suggest that *Search* involves trying to find unknown forms or meanings by consultation, e.g. a dictionary or teacher. In our case, the learners may ask the system for help, or each other, as in Extract 1 below. *Evaluation* is a cognitive factor looking at how learners employ a word to establish its relationship with other words. When using the kitchen, learners may employ the words in such a way when speaking to each other during or after the main task, but they are not obliged to do so. Recording of learner use shows that some learners do so, but others do not. So we can conclude that, in terms of task-induced involvement, the kitchen task is strong in terms of Need and

Search but Evaluation is variable, since it is dependent on the learners' inclinations rather than on the task or system. The extract below, taken from the English sub-corpus, illustrates how these concepts relate to users' practices.

**Extract 1**

```
1     KIT:  next (.) sift the flour into the mixing bowl
2     L1:   [six
3     L2:   [sift flour well I don't know what sift means so shall
4                                  [we request help
5     KIT:                      [((indicates that HELP 1 is available
6           by flashing))
7     L1:   yes
8     L2:   ((shakes red help tool))
9     KIT:  next (.) sift the flour into the mixing bowl ((provides HELP
10          1 for during-task))
11    L1:   mixing bowl
12    L2:   okay maybe sift here
13    L1:   yeah
14    L2:   ((pours all the flour directly into the mixing bowl))
15    KIT:  ((indicates that HELP 2 is available indication))
16    L2:   ((shakes green interaction tool))
```

```
17    L2:   ((selects HELP 2 for during task through the GUI))
18    L1&2: ((both look at the screen which provides HELP 2 showing picture
19          of sifting process with accompanying text))
20    L2:   oh:
```



```
21    L2:   ((Scoops flour with the sieve)) I can sift it now do you
22          think it will be the same
23          ((Flour falls on counter))
24    L1:   oh my god
25    L2:   okay ((pours back into the weighing scale then begins to
26          sift)) do you think it makes a difference
27    L2:   maybe yes
28    L1:   it's better er: I don't know
29    L2:   it's thinner maybe
```

In line 3 we see the learners noticing a *need* to understand 'sift' and they employ a *search* strategy by requesting help. The HELP1 audio repetition (line 9) does not enlighten then, whereas the photo of the process in HELP2 (lines 18 and 19) does do so. We can see the *evaluation* stage in lines 21-29 in that L2 retrieves the new word and uses it creatively 'I can sift it now'. Moreover, the new word is related to real-world cooking concepts in that the students wonder (lines 21-29) whether sifting the flour makes any difference to the product. Their conclusion in line 29 is that the flour may be thinner as a result of sifting. The users orient to the system in a similar way as they might to a teacher in TBLT, namely as a co-interactant or a resource which can aid them in the resolution of task-related and interactional trouble.

**Two studies on vocabulary learning in the EDK**

In order to assess the effectiveness of the EDK's cooking sessions in promoting vocabulary learning, two studies were set up by the English and the Italian research teams. We chose vocabulary acquisition as a focus since the task involved identifying and physically manipulating a set of concrete kitchen utensils and food ingredients, which formed a cohesive vocabulary set. Moreover, L2 vocabulary assessment is a mature research field (Read 2000) with established procedures and tools. The assessment structure for active verbal production of lexical items by individual learners was that of pre-test, post-test and – in Study 1 only – delayed post-test. The

assessment design and processes were 'wrapped around' the task design and processes so that there would be minimal impact on the holistic, autonomous nature of the task itself (see Figure 1 above).

In both studies, the aim was to learn a specialised lexical set related to cooking. In this test users had to produce the spoken form of words in connection with physical objects which are employed in a cooking task. A situational set (Nattinger 1988: 72) of 16 items of cooking utensils and ingredients was chosen, all of which were nouns. The purpose of the test was to determine the extent to which an individual user's verbal production of specific vocabulary items increased as a result of a cooking session. The types of word knowledge being tested here were connecting spoken form to meaning (Nation, 2001: 47) in relation to concrete objects, and production of phonological form.

In a comparable study, R. Ellis and He (1999) studied the acquisition of a lexical set (furniture) of 10 items in an experimental study of the effects of premodified input, interactionally modified input and modified output. Scoring options were either 0 or 1 for each lexical item. In our study, we wanted greater granularity in relation to spoken production of individual lexical items, since 'A sensitive vocabulary test can show that there has actually been small amounts of learning from some low-strength teaching or learning intervention.' (Nation and Webb 2011: 304). We therefore adapted Barcroft's (2002) proposal for the Lexical Production Scoring Protocol-Written (LPSP-Written), which quantified the ability to spell words. The following scheme was created for the quantification of the oral production of specific individual lexical items (Figure 5). The marking system was trialled with participants before data collection and inter-rater reliability of 96% was achieved.

**[FIGURE 5 NEAR HERE]**

## STUDY 1

The first study was conducted on 25 pairs of learners of L2 English, following the EDK's directions for preparing scones. The cooking sessions (lasting about one hour) involved different pairings based on learners' level of ability in English and cooking skills. We tried to pair participants so that one had a higher level of English and the other was better at cooking. This then created an information gap and potential for information transfer between participants. Which languages were spoken by participants when using the kitchen? In many cases the two participants did not have a common L1 and spoke L2 English the whole time. In some cases the participants had a common L1, but spoke English L2 all of the time, whereas in other cases they had a common L1 and spoke a mixture of L1 and English L2. There were 50 learners of L2 English resident in the UK, who were assessed on the same 16 vocabulary items (utensils and ingredients) on 3 occasions and with 5 rating options. Self-reported

information on the participants is as follows. They came from a range of L1 backgrounds including Chinese (n=31), Arabic (n=3), Korean (n=2), Spanish (n=4), Thai (n=5) and Turkish (n=2). Ages ranged from 20-58 with an average age of 27. There were 43 females and 7 males in total. Length of time spent learning English ranged from 1-35 years with an average of 11 years.

We delivered the assessment in the 3-stage test cycle as follows (figure 1). We chose 16 vocabulary items (kitchen equipment and ingredients) which the learners would use in the recipe and tested each individual of a cohort of 50 learners separately on them in the pre-test. We showed the testee each object and asked its name in L2 English, using an audio recorder to record what they said, if anything, for each item. We therefore established the extent to which each individual was able to actively produce each item prior to the cooking session, using the rating scale in figure 5 above. After they finished the cooking task, each individual completed the post-test separately following exactly the same procedure as the pre-test. We assessed them again individually on exactly the same items two weeks later as a delayed post-test. We were therefore able to record granular evidence in terms of individual changes in active production of the specific vocabulary items over a period of two weeks.
The results are presented below. In all cases, repeated measures ANOVA was used to investigate the significance of gain between pre, post and delayed post-tests.

**[FIGURE 6 NEAR HERE]**

Figure 6 and Table 1 show that the mean score for an aggregate of all items and all participants rose from 8.51 in the pre-test to 12.24 in the post-test. These differences were all statistically significant (see table 2).

**[TABLE 1 NEAR HERE]**
**[TABLE 2 NEAR HERE]**

Figure 6 shows an aggregated mean score for all items, but there was considerable variation in the learning of individual vocabulary items. Two contrasting cases are shown in figure 7 below.

**[FIGURE 7 NEAR HERE]**

Here we see that the mean score for ‘sieve’ for all 50 learners was only 0.09 in the pre-test, rising to 0.55 in the post-test, a striking rise of 0.46 with a significance level of .0001. By contrast, the well-known item ‘milk’ already had a mean score of 0.98 in the pre-test, rising to 1.0 in the post-test, a gain of only 0.02. So we can conclude that there was a significant gain in the mean score between pre-test and post-test for these items when aggregated. However, the degree of gain for individual items showed considerable variation; there is a prima facie case that

this variation was related to the degree of prior knowledge of the vocabulary item, although other influences cannot be excluded.

Figure 6 also showed that there was a further increase from the post-test (12.24) to the delayed post-test 2 weeks later (13.17). Post-hoc tests revealed a significance of .004. This increase was entirely unexpected and therefore deemed worthy of further investigation. In the 2-week period between the post-test and the delayed post-test, the learners were immersed in an English-speaking environment and hence it is possible that they encountered the newly-learned items or looked them up. They had, however, been given no instructions at all about what they should (or should not) do during the 2-week period. We therefore sent a questionnaire to all learners and received 17 responses. Figure 8 below shows the percentage of responses in each category.

**[FIGURE 8 NEAR HERE]**

Whilst a relatively small number of students did not do anything, the majority encountered the vocabulary items again in their environment and some students employed active learning strategies in relation to the items. Furthermore, the test and questionnaire results suggest that some of the learners perceived their learning experience in the pervasive digital environment as having relevance to their real-world lives and also that they felt motivated to continue to work on the same vocabulary learning in their real-world lives.

## STUDY 2

Study 2, on Italian, aimed to answer the same research question as Study 1, viz. 'do cooking sessions in the EDK provide opportunities for language learning, as operationalised in terms of vocabulary knowledge?', and it followed a similar research design. Participants' productive knowledge of vocabulary items was tested before and after the session, with the same type of test-task integration described in previous sections. Participants' initial competences were comparable in the two groups, as their scores in the pre-test were exactly the same, i.e. 5.3 points out of a theoretical maximum of 10.

There were however some methodological differences. First, there was no delayed post-test in Study 2, as it turned out to be practically impossible to interview participants for a third time. Secondly, participants to Study 2 did not cook all the same recipe, as 20 prepared spaghetti with tomato and olives (Recipe 1) and 14 meat rolls in tomato sauce (Recipe 2). Thirdly, the pre-task in the English cooking sessions involved picking up all the ingredients and the utensils to be used in the recipe, while in the Italian ones participants were asked to gather the ingredients only, to make the task shorter and lighter. Finally, and most importantly, for both recipes in Study 2 the pre- and post-test did not involve the same set of lexical items, as was the case in Study 1, but two sets of

completely different ones. Assessing the same items in all three testing moments has the clear advantage of allowing one to compare how learners' knowledge of these items varies over time and thus, arguably, because of their participation to the cooking sessions. However, a possible objection to this approach could be that some of the observed variance might be due to a 'test-retest' effect, i.e. that learners may have learned some new vocabulary *by taking the test*, in addition to their having been involved in the main cooking task. In order to control for this effect, Study 2 used completely different sets of words in the pre- and post-test. The sets were generated using random numbers and their administration was counterbalanced, so that half of the participants received set A in the pre-test and set B in the post-test, and the other half the reverse, which allowed us to control for any possible bias introduced by one of the two sets being more or less difficult than the other. Given that two different sets of words were created for the pre- and the post-test, they contained 10 words each, instead of 16 as in Study 1.

The 34 participants were 9 males and 25 females; their mean age was 46.3 (range 20 - 77) and their exposure to Italian varied between 1 week and 15 years, with a mean of 2:8 years. Some of them were long-term residents in Italy, others were there to attend language courses lasting from a few weeks to several months. According to self-reported information we collected, they came from a range of L1 backgrounds including German (n=8), English (n=5), Danish (n=3), Finnish (n=2), French (n=2) and Albanian (n=2) and other European and non-European languages. In many cases, the two participants did not have a common L1 and spoke L2 Italian most of the time. Only in few cases did they share the same L1 (be it Danish, English, Finnish or German) and spoke a mixture of L1 and Italian L2.

The Italian cooking sessions followed the same protocol as the English ones, but while the latter took place in a real kitchen, the former were organised in a variety of settings (from house kitchens to school classrooms), using a portable hotplate. Irrespective of the different contexts, all the sessions included the same steps. While one of the two learners was asked to go out of the room and start filling in the biographical information questionnaire, the other was involved in the pre-test, which consisted in naming 10 objects chosen among the ingredients and the utensils of the recipe. Each item was shown to the learner and his/her reactions were audio/video recorded. Every 5 seconds, the Research Assistant moved on to the following item. The pre-task consisted in collecting the ingredients needed to make the recipe, as is often the case on cooking-themed programmes on TV. The main task was the cooking activity proper, following the kitchen's instructions step by step. After the cooking activity, users were generally very eager to taste the dish they had just prepared. While doing so, they took part in short unstructured interviews reporting on what they had learnt, which could concern cooking, language, technology or all of these; all interviews were video recorded. Participants finally took the post-test aiming at assessing their individual skills after the cooking session. While one of the two learners was asked to leave the room, the other was shown 10 items chosen among the recipe's ingredients and utensils, not included in the pre-test, and had to name each of them.

All words in the pre- and post-test lists appeared in the input provided by the EDK and were thus potentially relevant for the cooking tasks, with the exception of one word included in the tests for Recipe 2, *grembiule* (apron), which did not occur in the cooking session. Values for this item were thus discarded from quantitative analysis.

Tables 3 and 4 present the descriptive statistics for the participants who cooked Recipe 1 and 2, respectively, while Table 5 reports results for the whole group. Scores in the post-test were higher for both Recipe 1 and Recipe 2, thus showing that participation in a one-hour cooking session can have a positive effect on the acquisition of task-specific vocabulary. The differences were not statistically significant in a paired-samples t-test, probably because of the small size of the samples: for Recipe 1, t = -1.8071, df = 19, p = 0.087; for Recipe 2, t = -0.99404, df = 13, p = 0.34. However, the t-test applied to the whole group of 34 participants produced a p-value extremely close to statistical significance, t = -2.0049, df = 33, p-value = 0.053).

**[TABLE 3, 4, 5 NEAR HERE]**

The group score increased between pre- and post-test by +15.5% for Recipe 1, by 13.5% for Recipe 2 and by +15.3% for the two groups combined. The gain was much larger in Study 1, where learners' scores increased by +43.9% in the immediate post-test and +54.8% in the delayed post-test. This difference may be due, among other things, to the fact that the pre-task in Study 1 was longer and involved all ingredients *and utensils*, while in Study 2 only the ingredients' names were pronounced by the kitchen. Also, participants in Study 1 were made aware of some of their lexical gaps in the pre-test, which might have led them to pay extra attention to these unknown words during the main task, favouring their acquisition and thus improving performance on these very items in the post-test. Such an effect was intentionally removed from Study 2, where items in the pre- and in the post-test were different. However, the clear and consistent gains achieved in Study 2, too, show that learners' culinary vocabulary does increase after a cooking session, even in the absence of a test-retest effect.

As in Study 1, there was a fair amount of variation among learners, as evidenced by the relatively high standard deviations, but also among different lexical items. Figures 9 and 10 display the scores for individual words before and after the cooking sessions, for Recipe 1 and 2, ranked from lower to higher gains. It should once again be emphasised that the same word appeared for some participants in the pre-test and for others in the post-test: the fact that for some items the average score was lower in the post-test should thus not be taken to mean that some learners 'unlearned' these words, but simply that the average for the group who took the post-test was lower than that of those who took the pre-test.

**[FIGURE 9 AND 10 ABOUT HERE]**

In the first recipe, pasta with tomato, olives and capers, many words were known by most participants even before the task, beginning, unsurprisingly, with *pasta*, which was known by everybody. The words that appeared to be most problematic were *lid, drainer, scale, chopping board, stove*, whose average score, except for *lid*, was in any case higher in the post-test than in the pre-test. These differences were even more dramatic for the second recipe, meat rolls in tomato sauce. Some words, like *meat mallet* and *spatula*, were completely unknown to all of those taking the pre-test; others, like *chopping board* and *capers*, were barely attempted by two participants. These are also the words achieving the biggest gains in the post-test, where several participants showed fair or even full knowledge of them.

**Conclusions**

Because digital pervasive environments for language learning as described here are intended to provide a holistic learning experience, their overall contribution to language learning would best be portrayed and evaluated in a holistic manner, as in chapter 3. Nonetheless, this chapter has shown that it is possible to adapt the environment to focus on one specific aspect of language learning (in this case, vocabulary) and to provide targeted input for learning. Furthermore, it is possible to design an assessment structure to provide evidence of learning of the targeted feature which does not interfere with the smooth running of the pervasive environment, if this is 'wrapped around' the task cycle (see figure 1) in which the main activity (in this case, cooking) takes place. Also, the test cycle was integrated, especially in Study 1, into the immersive experience, enabling participants to perceive a need to learn specific items and showing them which they had (or had not) learnt. Furthermore, the questionnaire in Study 1 revealed that some learners were able to apply the vocabulary they had learnt in the EDK to the real world outside, without any prompting to do so.

In relation to vocabulary learning, the two studies have shown the following. From a methodological point of view, the two protocols proved complementary and equally useful in tracking the extent and progress of participants' learning. The procedure followed in Study 1 allows one to show how specific lexical items are gradually mastered by single individuals and by the whole cohort; it also demonstrates that a pre-task phase in which participants are made aware of their lexical gaps can foster subsequent learning by promoting focus on form during task performance, and the same effect can be observed in the period between the post-test and the delayed post-test. The procedure followed in Study 2 neutralises test-retest effects, including drawing participants' attention to specific lexical items.

However, even under these stricter conditions, participants' lexical knowledge was consistently higher in the post-test, thus showing that merely participating in the cooking session does produce measurable gains after

about one hour on task. Study 1 additionally demonstrates that a testing phase included in the task cycle may further boost learning by promoting incidental focus on form on unknown or partially known vocabulary items, which has clear pedagogic implications and opens up fruitful research perspectives on the virtuous relationships between formative assessment and learning in a pervasive digital environment.

From a theoretical point of view, it is unclear precisely what factors have enabled learning and further research is required to disaggregate these. This was a kinesic, multimodal, task-based experience in a specific real-world context, learning a lexical set of concrete words, specifically nouns. The objects were physically manipulated and all of the senses were engaged. Testing procedures closely resembled learning procedures and took place in the same location. According to the studies cited in the chapter's first sections, all of these factors may have contributed to making the words relatively easy to learn, although the requirement for active spoken production made the test relatively challenging.

The Newcastle team has therefore undertaken a further research project (The Korean Digital Kitchen) in relation to learning of the Korean language, culture and cuisine to investigate how each of these factors plays a specific role. Using a quasi-experimental research design, students learnt Korean vocabulary a) in the same way described in this book b) by merely looking at photographs of kitchen objects. The quantitative data showed that students learn significantly better when touching and manipulating the utensils and objects as part of the cooking task than when looking at photographs; a publication is in preparation.

**References**


Allen, E. and R. Vallette (1972), *Modern Language Classroom Techniques: A Handbook.* Harcourt, Brace and Jovanovich Inc.

Barcroft, J. (2002), 'Semantic and structural elaboration in L2 lexical acquisition.' *Language Learning* 52 (2): 323-363.

De Groot, A. (2006), 'Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting', *Language Learning*, 56 (3): 463-506.

Ellis, N. and A. Beaton. (1993), 'Psycholinguistic determinants of foreign language vocabulary learning.' *Language Learning,* 43 (4): 559-617.

Ellis, R. (2012), *Language teaching research and language pedagogy,* Chichester (UK): John Wiley & Sons.

Ellis, R. and X. He (1999), 'The roles of modified input and output in the incidental acquisition of word meanings.' *Studies in Second Language Acquisition,* 21: 285-301.

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., and F. Binkofski (2010), 'Grasping language – A short story on embodiment', *Consciousness and cognition*, 19 (3): 711-720.

Lakoff, G. and M. Johnson (1999), *Philosophy In The Flesh: the Embodied Mind and its Challenge to Western Thought*, New York: Basic Books.

Lantolf, J.P. (2000), 'Introducing sociocultural theory', in J.P. Lantolf (ed), *Sociocultural theory and second language learning*, 1-26, Oxford: Oxford University Press.

Laufer, B. (1997), 'What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition', In Schmitt and McCarthy (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge University Press.

Laufer, B. and Z. Goldstein (2004), 'Testing vocabulary knowledge: Size, strength, and computer adaptiveness', *Language Learning*, 54 (3): 399-436.

Laufer, B. and J. Hulstijn (2001), 'Incidental vocabulary acquisition in a second language: the construct of task-induced involvement.' *Applied Linguistics*, 22 (1): 1-26.

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

Melka, F. (1997), 'Receptive vs. productive aspects of vocabulary'. In Schmitt and McCarthy (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge University Press.

Nation, I. S. (2001), *Learning vocabulary in another language*, Cambridge: Cambridge University Press.

Nation, P. and S. Webb, (2011), *Researching and Analyzing Vocabulary*. Boston: Heinle.

Nattinger, J. (1988), 'Some current trends in vocabulary teaching'. In Carter and McCarthy (eds.) *Vocabulary and Language Teaching*. Harlow: Longman.

Oxford, R. and R. Scarcella**.** (1994), 'Second language vocabulary learning among adults: State of the art in vocabulary instruction'. *System,* 22(2): 231-243.

Read, J. (2000), *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N. (2010), *Researching vocabulary: A vocabulary research manual*, Basingstoke, UK: Palgrave Macmillan.

Schmitt, N. and M. McCarthy (eds.) (1997), *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge University Press.

Vygotsky, L. S. ([1934] 1986), *Thought and Language*, Cambridge, Mass: MIT Press.


Figure 1 The test cycle (white) and the task cycle (blue)

Figure 2: Red and green Interaction Tools


Figure 3: HELP 2:  multimedia gloss in the Pre-Task phase


Figure 4: HELP 3 multimedia gloss video in the During-Task phase

| | |
|---|---|
| 0.00 points | The speaker says nothing at all or states that s/he is unable to answer. |
| 0.25 points | The speaker makes an attempt to name the target object which is unintelligible and is very difficult to understand in relation to the target object. |
| 0.50 points | The speaker produces the target lexical item partially, or in a way which can only be understood to relate to the target object with some difficulty, with a major problem in pronunciation and/or clarity. Or the speaker tries to describe the object rather than name it. |
| 0.75 points | The speaker produces the entire target lexical item in an intelligible way, but with a minor problem in pronunciation and/or clarity, or in delivery. |
| 1 point | The speaker produces the entire target lexical item with precision and clarity. |

Figure 5. Lexical Production Scoring Protocol (Spoken)

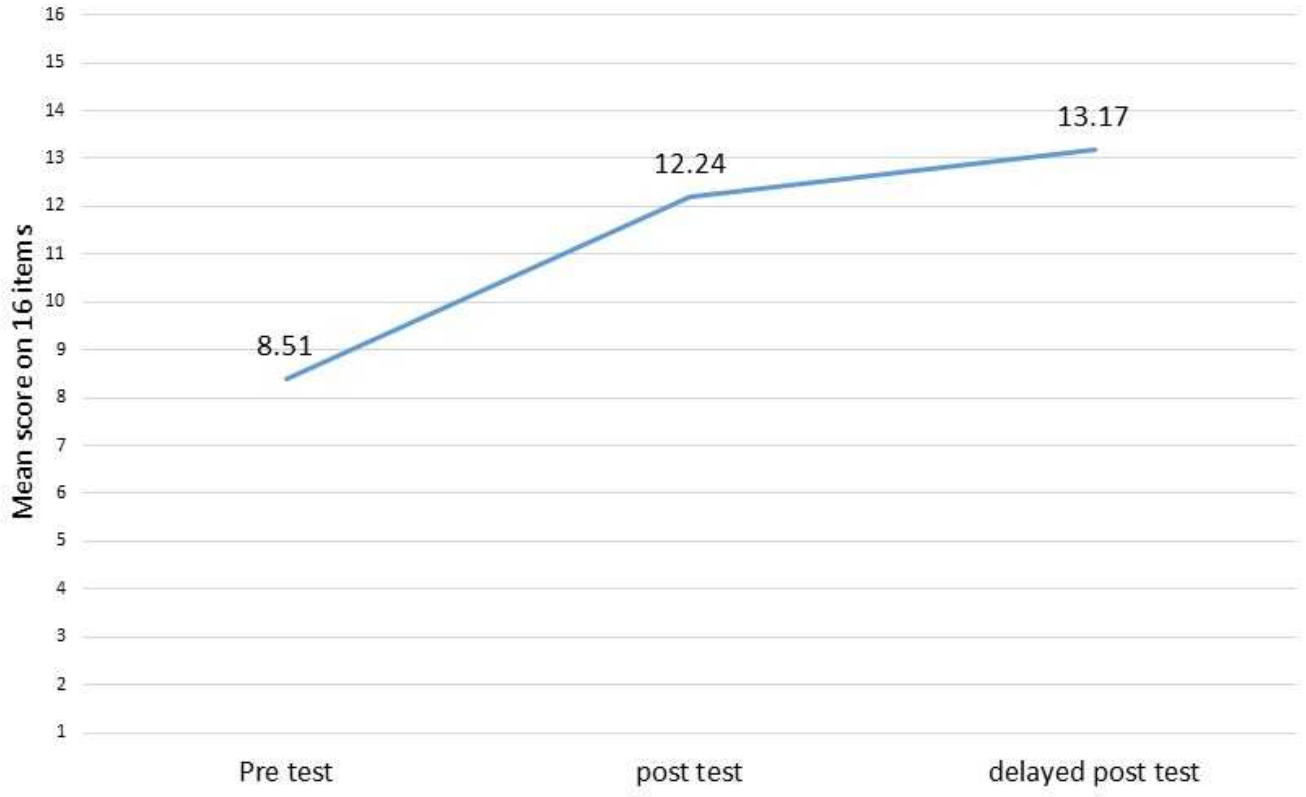



Figure 6: Study 1, mean score on 16 items in pre, post and delayed-post test


Figure 7: Scores for ‘sieve’ and ‘milk’ in pre, post and delayed post-test

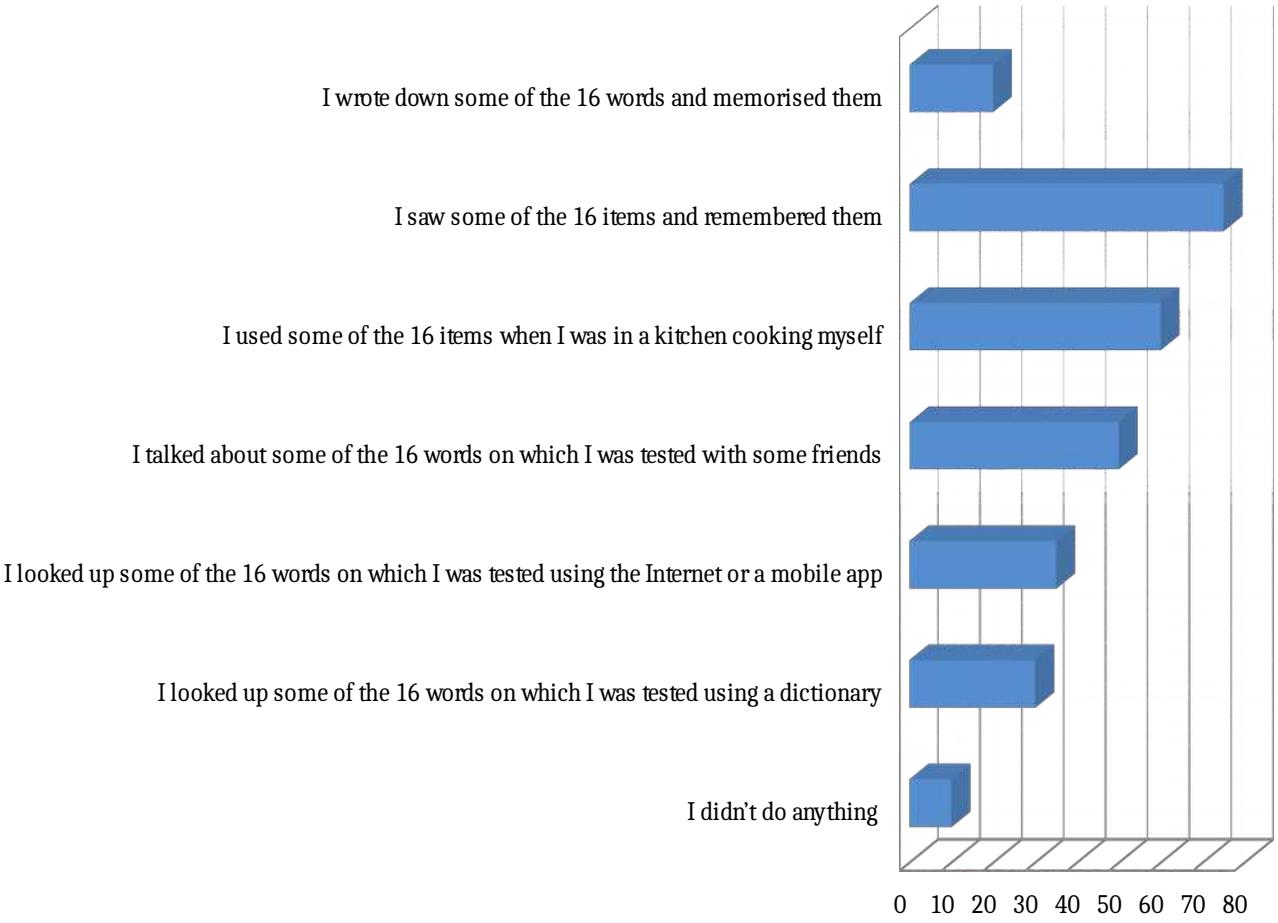



Figure 8: Activities undertaken by learners between the post and delayed-post test

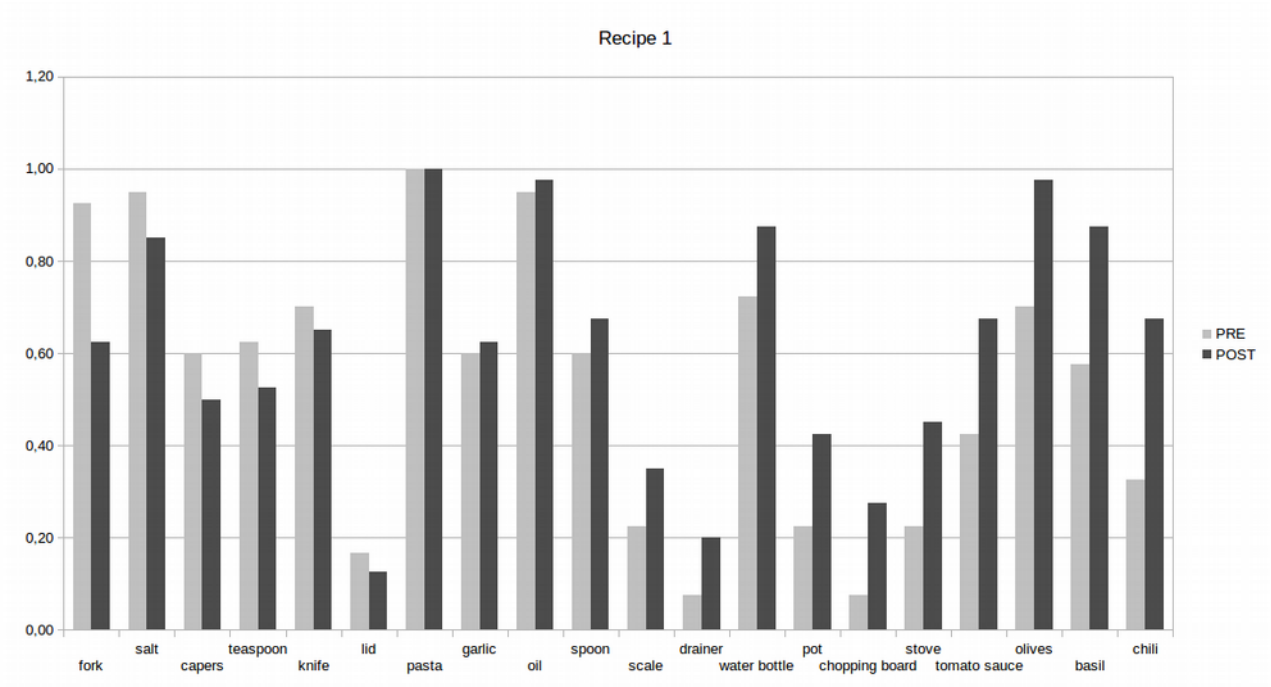



Figure 9: Study 2, pre- and post-test scores for individual lexical items in Recipe 1.

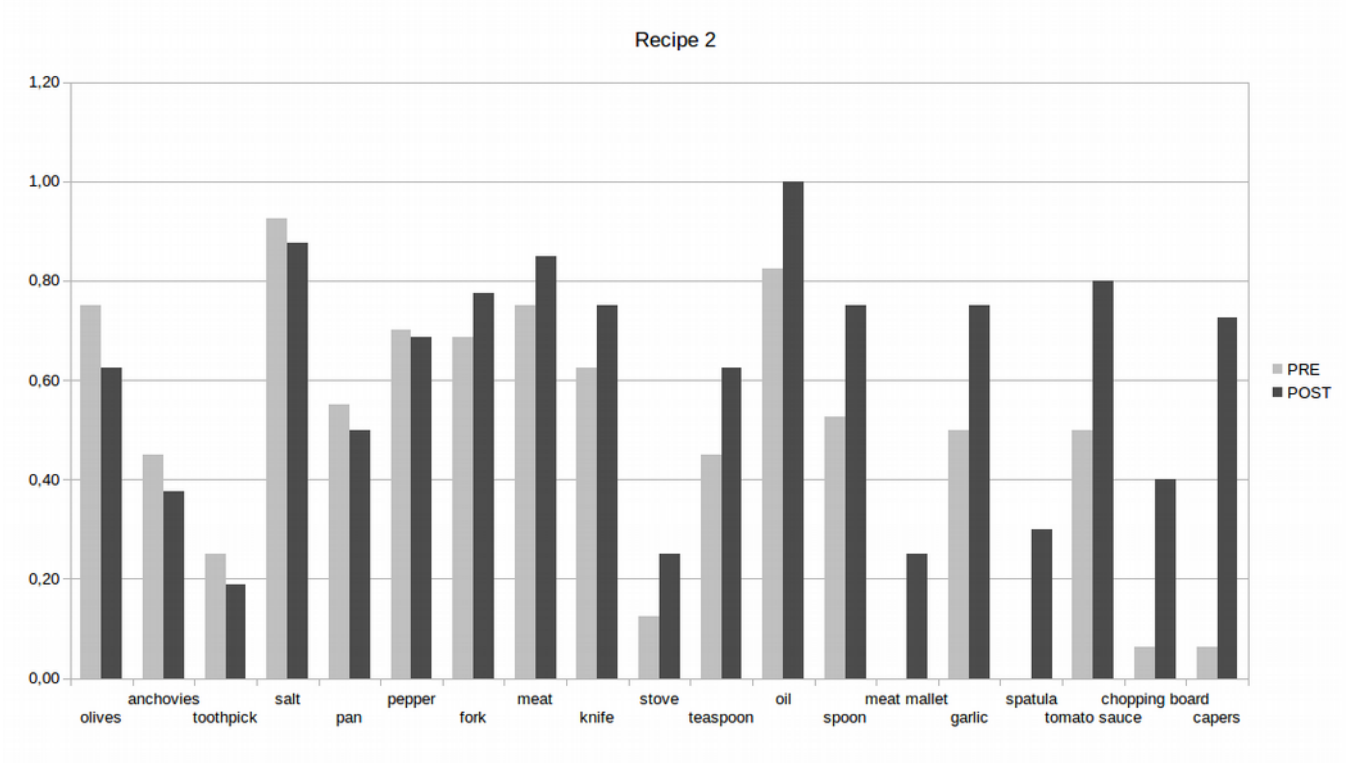



Figure 10: Study 2, pre- and post-test scores for individual lexical items in Recipe 1.

| **Overall score** | Mean | Std. dev. |
|---|---|---|
| Pre test | 8.51 | 1.58 |
| Post test | 12.24 | 2.09 |
| Delayed post test | 13.17 | 2.02 |

Table 1. Study 1, Mean and standard deviation scores

| **Post-hoc tests *overall*** | | *Sig* |
|---|---|---|
| Pre-test | Post test | .0001 |
| Pre-test | Delayed post test | .0001 |
| Post test | Delayed post test | .004 |

Table 2. Study 1, statistical significance in relation to the tests

| **Overall score** | Mean | Std. dev. | N |
|---|---|---|---|
| Pre test | 5.30 | 1.91 | 20 |
| Post test | 6.12 | 2.39 | 20 |

Table 3. Study 2, Recipe 1, Mean and standard deviation scores

| **Overall score** | Mean | Std. dev. | N |
|---|---|---|---|
| Pre test | 5.17 | 2.87 | 14 |
| Post test | 5.87 | 1.85 | 14 |

Table 4. Study 2, Recipe 2, Mean and standard deviation scores

| **Overall score** | Mean | Std. dev. | N |
|---|---|---|---|
| Pre test | 5.24 | 2.31 | 34 |
| Post test | 6.04 | 2.16 | 34 |

Table 5. Study 2, all participants, Mean and standard deviation scores