DOI: 10.20535/2410-8286.231421

PROCEDURES FOR ASSESSING COGNITIVE SKILLS OF PROSPECTIVE LANGUAGE TEACHERS

Iryna Sverdlova
V.N. Karazin Kharkiv National University, Kharkiv, Ukraine
iryna.sverdlova@karazin.ua

The purpose of the research was to find out how the procedures for measuring students' cognitive skills could be incorporated into the university course Language Teaching Methodology. The study was organised within a framework of Anderson's theory of cognitive skills development and Glaser's taxonomy of dimensions for assessing achievement. We developed the instrument, which encompassed two empirical questionnaires for treatment groups. Both questionnaires comprised an equal number of tasks but differed in the content of procedures for measuring knowledge acquisition and structure as one of the dimensions of cognitive assessment. Empirical Questionnaire 1, based on a traditional approach to assessment, included multiple-choice questions related to the lecture material. Empirical Questionnaire 2 comprised both traditional and unconventional measures, such as a SVT test; constructed and conversation-based responses; simple and high order rule tasks. Thirty-four third-year students of the Department of Foreign Languages, V.N. Karazin Kharkiv National University participated in three-stage research-oriented teaching, which lasted ten weeks. Both qualitative and quantitative data analysis methods were employed for the evaluation of learning outcomes. After final testing, we compared the results obtained from the students. The group mean difference (EG vs. CG) was 0.12 points, 95% confidence interval (0.07 - 0.17), two-sample t-test $p < 0.0001$. Findings suggest that cognitive skills assessment considerably affects and improves student learning. The implications relate to final grades assessment and curriculum design and contribute to expanded uses for cognitive skills testing.

Keywords: cognitive skills development; taxonomy of dimensions; assessment; measure procedures; prospective language teachers.

Introduction

Assessment is a critical component of the courses where there is a strong relationship between theory and practice, Language Teaching Methodology in particular. Schools are placing greater demands on institutions of higher pedagogical education, and knowledge about students' learning outcomes is becoming of crucial importance. The ability to provide sensible measures of professional skills is under increased scrutiny. However, there is a lack of information on assessing cognitive skills, which are important for career advancement. Hence, the pivot question is what procedures can be used to assess progress in the course designed to enhance both the professional and cognitive skills of prospective language teachers.

In recent years, research has consistently shown that effective formative assessment techniques can and do improve student achievement and learning (Jönsson, 2020; Voinea, 2018; Tigelaar & Sins, 2020; Wang, Sun & Jiang, 2018).

According to O'Keeffe et al., (2020); and Cole (2010), formative assessment is used to conduct in-process evaluations; its goal is to provide initial feedback to the students and educators. Instructors need sensible measures in order to 1) design courses that promote the development of a sound on-the-job educational environment; 2) assess the need for filling gaps in students' knowledge within the teaching period; 3) select the best methods for enabling students to go beyond knowledge-level cognitive operations to achieve best academic performance. Students need a regular evaluation of knowledge because it 1) helps to realise graduate attributes; 2) provides with a solid guide for studies; 3) assists in preparation for the course assessment.

Assessment of students' learning outcomes from a cognitive perspective has been a hot topic of research in education since 1985, when the Buros-Nebraska Symposium first addressed the issue of the influence of cognitive psychology on testing skills (Benton & Kiewra, 1987). Recognising the challenges facing the new school, the Ministry of Education and Science of Ukraine released The Expert Assessment of Professional Action Competence – the strategic document which encourages the assessment of teachers' both activity and cognitive competencies.

Some theory developers assert that cognitive competency is more than an ability to manipulate and strategise information, but an ability to internalise, self-regulate and transfer cognitive skills to construct knowledge and make sense of the surrounding (Piaget (1977); Vygotsky, 1962; Vygotsky, 1978). A cognitive skill refers to a person ability to gain meaning from experiences and information (Ainsworth, 2013), it is a product of learning (Ullah et al., 2019), and it has a distinctive history of quantitative and

© Iryna Sverdlova. 2021. Published by Igor Sikorsky Kyiv Polytechnic Institute. This is an Open Access article distributed under the terms of the licence CC BY 4.0

qualitative developmental change (Royer, Cisero & Carlo, 1993). "Cognitive skill development can be viewed as a gradual process of transition through the three hierarchical layers: 1) a layer of basic capacities (memory capacity, speed of concept activation); 2) a layer of cognitive skills that are capable of being transformed from controlled to automatic/encapsulated processes; 3) a layer of higher cognitive skills and capacities that are responsible for goal setting and planning of cognitive activity" (Royer, Cisero & Carlo, 1993, p. 204). The middle and upper layers are of special interest for our course.

Skill acquisition is a three-stage process (Anderson, 1982). At the first, the declarative stage, a learner can answer questions about the skill, and he/she can perform the skill by interpretatively utilising declarative information. At the second, the knowledge compilation stage, the information acquired in the declarative stage is transformed into a procedural form that can be applied with minimal conscious reasoning activity. At the third, the procedural stage, the newly acquired productions become strengthened, their conditions for execution are more completely specified, and considerable learning entails the speeding up of a particular skill application. Cognitive skills are applicable to a number of activities within a defined domain of activity, but their use is generally confined to that domain (Royer, Cisero & Carlo, 1993).

The cognitive approach to assessment skills suggests that the following factors contribute to successful academic behaviours: the learner's declarative knowledge, procedural knowledge, control processes, cognitive strategies, and metacognitive processes (Benton & Kiewra, 1987). Hence, the first discussion of the procedures for assessing cognitive skills can be organised around a framework of declarative knowledge (which refers to knowledge of facts and information) and procedural knowledge (which refers to the knowledge of how to perform a specific task).

For describing skill assessment procedures Glaser, Lesgold, and Lajoie (1987) provide a taxonomy of dimensions, which is described as a set of components common to developing skills. These dimensions include: 1) knowledge organisation and structure; 2) depth of problem representation; 3) quality of mental models; 4) efficiency of procedures; 5) automaticity to reduce attentional demands; 6) proceduralised knowledge; 7) procedures for theory change; 8) metacognitive skills for learning.

Over the past decades, significant progress has been made on theoretical and applied aspects of various procedures for assessing cognitive skills. Some studies have focused on measuring declarative and procedural knowledge in different domain of study. Richter-Beuschel, Grass and Bogeholz (2018) dwell on measuring procedural knowledge for solving biodiversity and climate change challenges within a framework of science courses. McIlwain and Sutton (2015) develop methods for measuring breadth and depth of knowledge in sporting environments and argue that experts and novices represent problems in different ways. Vandierendonck (2017) investigates the dimension of atomicity of performance; describes the technique that involves measuring speed and accuracy simultaneously, and provides an estimate of the resource load that accompanies task performance. Ismail (2016) reports on developing semiotic declarative knowledge models about magnetism for prospective science teachers. Boruff and Harrison (2018) analyse how knowledge and skills are assessed in the information literacy (IL) instruction for rehabilitation sciences students. Lenz et al., (2020) develop a test instrument that affords valid measurement of students' conceptual and procedural fraction knowledge.

Royer, Cisero and Carlo (1993) state: "The overview of measurement procedures that many of the assessment techniques can be used with benefit during each of the stages of skill development" (p. 208).

The purpose of this article is to present the procedures for measuring knowledge acquisition, organisation and structure, which look highly promising for the Language Teaching Methodology course. The following three questions were addressed.

1. Question 1. What procedures could be used to assess students' learning outcomes from a cognitive perspective?
2. Question 2. Does a relationship exist between formative assessment and students' acquisition of lecture material?
3. Question 3. Is there a relationship between the measuring procedures' content and students' performance?

To achieve the purpose, the objectives were set: 1) to select appropriate measurement procedures for the assessment of students' cognitive skills; 2) to carry out a formative assessment; 3) to assess the efficacy of the suggested procedures.

The hypothesis is stated in the following form. The incorporation of procedures for measuring knowledge acquisition into the formative assessment of the course Language Teaching Methodology will allow students to develop strong cognitive skills and improve their learning outcomes in the course. Our conjecture is that further study of the relationship between formative assessment, and student performance may provide additional information on testing cognitive skills.

Method

Research design

The study involved the use of mixed (quantitative and qualitative) data assessment. Qualitative methods included pedagogical observations; the analysis of data obtained via empirical questionnaires; survey, and focus groups (group discussions). Quantitative methods were applied to verify if the hypothesis was true. The obtained numerical data were analysed using mathematical and statistical methods, such as Fisher's exact test; a paired t-test; and a two-sample t-test.

Participants

The participants were the third-year students of the Department of Foreign Languages, V. N. Karazin Kharkiv National University, who took a course on Language Teaching Methodology in the 6th semester. In January 2021 we invited 40 students, based on a stratified random selection to collaborate in the research. After discussions 34 students gave a positive response to the invitation, six students made the choice not to participate. Therefore, we obtained the consent of free and conscious decision for the participation in the research experiment and the obtained data processing from 34 students. The students could withdraw without any consequences on their status. The researchers (2 in number) acted as instructors, assessors and reflectors, focusing on lecturing, giving complete instructions, answering students' questions, judging the tests responses, and analysing the outcome results.

Instruments and procedure

Research-oriented teaching began three weeks after starting the semester and covered 20 hours (2 hours per week). It was divided into three stages: (1) diagnostic (entry testing), (2) empirical training (questionnaires implementation), (3) checkout (final testing). The participants were randomly divided into equal in number groups: a control group (CG) and an experimental group (EG). For gathering data on the study, the following instrument was developed: 1) an entry test, 2) two questionnaires for empirical training in CG and EG, 3) a final test.

Diagnostic stage. The aim of the entry test was to assess students' acquisition of declarative and procedural knowledge gained at the first three lectures. The test included 24 multiple-choice questions regarding the learning content, and it was distributed without any notice in the last hour of two consecutive teaching hours in the course. Hence, the students were unprepared for the task and the test was consequently approaching active knowledge. Before handing out the test, the students were shortly instructed (2–3 minutes). The response time was 40 minutes.

Empirical training stage. For empirical training eight questionnaires for control after each thematic unit were designed, the difference being the format for CG and EG. In CG empirical questionnaire there were 12 *wh*-questions (8 of them were multiple-choice questions, the rest – open-ended) related to the lecture material, while in the EG questionnaire the 12 questions/tasks were arranged in the research format described below.

An empirical questionnaire for EG comprised 12 questions with different levels of complexity. The questions were ranging from multiple-choice questions to open-ended: six – for measuring declarative knowledge (4 multiple-choice questions), the rest – for procedural knowledge (2 multiple-choice questions). In order to illustrate what we have done, we provide examples from the questionnaire.

For assessing declarative knowledge, which is fact-based, both multiple-choice questions and open-ended questions were used. There were two types of multiple-choice questions: a traditional test (*Question 1*) and a Sentence Verification Technique test (*Question 2*). *Question 1.* What is NOT speech activity? 1) reading; 2) translating; 3) listening.

A Sentence Verification Technique test (Royer, Cisero & Carlo, 1993) consisted of a passage, six sentences in length, and a task. The test addressed the question of whether a reader had understood a particular text, and was based on four types of text sentences. The first type (an original) was a copy of a sentence as it appeared in the passage. The second one (a paraphrase) was constructed by changing as many words as possible in an original sentence without altering the meaning of the sentence. The third type (a meaning change) entailed changing one or two words in the sentence so that the meaning of the sentence was altered. The final kind (a distractor) was a sentence that was consistent with the theme of the passage but was unrelated to any passage sentence. An examinee read the passage and then, in the absence of the text, judged each of the test sentences to be "yes" or "no" sentences. In our test, the first sentence was original. *Question 2*. Classify the following sentences into true or false. 1. Argument, that young children are better learners than others, is strongest when phonological features of a second language are considered. 2. There is evidence that young children are more able to acquire the phonological system of a foreign language than other people are. 3. Argument, that young children are better learners than others, is strongest when

morphological features of a second language are considered. 4. There is evidence that young children and others acquire the phonological system of a foreign language in the same way.

There were two kinds of open-ended questions: constructed and conversation-based responses (Jackson et al., 2018). Constructed responses allow students to express knowledge and skills through their own words, they can reduce the likelihood of guessing correct answers; but they also enable students to provide errant responses due to a lack of knowledge or a misunderstanding of the question (*Question 3*). Conversation-based assessments allow students to provide a more complete response and improve their score. Constructed responses were an obligatory component of an empirical questionnaire, conversation-based assessments were involved during focus groups (four in number) when test results were discussed. *Question 3.* What are the main types of reading? Which reading rate is most suitable for different types of reading? Why?

Designing questions for measuring procedural knowledge (knowing how), we took into account the hierarchy of procedural knowledge: discrimination, concepts, simple rules and high order rules (Gagne & Medsker, 1996). Discrimination questions refer to common knowledge; hence, they were excluded from the questionnaire. There were two concept close-ended questions, which required the ability to use definitions in deciding the choice, instead of just stating the facts like in the declarative multiple-choice questions. (*Question 4*). The technique used in the test was based on the idea that knowledge structure can be characterised by using indexes of associative memory. The procedure consisted of the following. Main concepts were obtained from a text. The students were asked to place randomly ordered concepts into the appropriate column, and their classification accuracy was compared to the issues from the text. *Question 4*. Classify the following into grammar-translation method or audiolingual method: 1) derived from classical method, 2) structural patterns are taught using repetitive drills, 3) no attention to pronunciation, 4) oral skills, 5) classes are taught in the mother tongue, 6) focus on reading, 7) is based on behaviourist theory, 8) material is presented in dialogue form.

Simple rule questions and high order rule questions were open-ended. Simple rule questions concerned the ability to apply the rules, not to state them. A Program (Benton & Kiewra, 1987) was used as an instrument, which made it possible to conclude if a learner was skilled at performing the activities involved in functioning in the domain (*Question 5*). The students were given a problem task and were asked to describe the steps to be taken to solve it. Then the learner's process model was compared to the model (*Guidance 1*) designed by the experimenter. The analysis of students' answers helped to pinpoint the procedural error(s) that he/she was making. *Question 5.* What is the order of types of exercises to form a flexible grammar skill? In 100 words, explain your reasons. *Guidance 1*: 1) comprehension, 2) imitation, 3) substitution, 4) transformation, 5) communication.

The outcome of high order rules tasks was that one could generate a new rule by combining old rules and use the new rule to perform a task (Bonner & Pennington, 1991). For the questionnaire, we have chosen a flowchart (Benton & Kiewra, 1987), which is a set of boxes and arrows used to represent the decisions one makes when solving a problem. Each of the items from the chart could be further subdivided into subitems to permit more detailed and task-specific issues to be addressed (Figure 1).

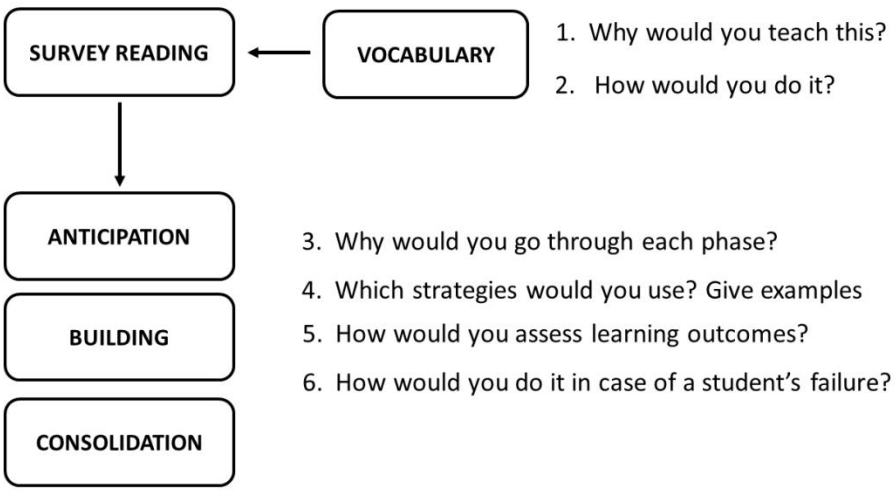

Figure 1: A flowchart for assessing high order rules tasks

A guidance (*Guidance 2*) for evaluating the responses was developed. *Guidance 2*. A student should demonstrate knowledge of basic methodological concepts regarding the following issues. 1. Vocabulary is the component skill that underpins second-language learners' text comprehension. 2. There is a certain order of the types of exercises and activities for developing vocabulary receptive skills. 3. Each stage of skills development is characterised by particular activities. 4. There are different approaches to assessing receptive skills and concrete tools for measuring learning outcomes. 5) Failures can be predicted, and there are certain efficient activities for overcoming the problem. A student can get from one to three scores for each question, depending on the quality of the answer.

After each thematic unit, the students of CG and EG got empirical questionnaires for self-regulated learning, and were to present the results of their work within a week. The instructor analysed the data and provided information to the students. In focus group discussions, the students of both groups could express their point of view by arguing about it.

Checkout. The aim of this stage was to check the effectiveness of an empirical questionnaires. The final test had the format of empirical questionnaire for EG and included 24 questions: 12 – for measuring declarative knowledge (6 multiple-choice questions), the rest – for procedural knowledge (6 multiple-choice questions). The response time was 40 minutes.

After the final exam, the students were surveyed about self-perceived readiness to answer different types of questions. We realised the importance of self-assessment interventions, which promotes students' use of learning strategies and effects on motivation and self-efficacy (Panadero, Jonsson & Botella, 2017). The students were asked to rate their level of readiness to answer different types of questions on a 5-level Likert scale: always (100%); very often (70% >); sometimes (< 50%); rarely (< 10%); never. They had to ponder the following questions. How often do you feel that you are ready: 1) to state facts; 2) to use definitions in deciding the choice; 3) to classify sentences into true or false; 4) to answer questions in written form; 5) to describe the steps to be taken to solve a problem; 6) to combine rules, suggesting an approach to problem solving; 6) to express your point of view orally.

Results

The issue we examined was whether the learning outcomes of the 34 students were affected by the assessment procedures used by researchers. We hypothesised that the incorporation of procedures for assessing cognitive skills into the course Language Teaching Methodology would allow students to improve their learning outcomes, i.e. students' declarative knowledge base would become more accurate, and students would be able to manage procedural questions. For measuring declarative knowledge at the checkout stage, a multiple-choice test, Sentence Verification Technique test, constructed and conversation-based responses were used. The procedural knowledge was decomposed according to the hierarchy proposed by Gagne (Gagne & Medsker, 1996). In this section, we provide the results obtained via research-oriented teaching.

At the diagnostic stage, we conducted the entry multiple-choice test (24 questions) to determine the amount of declarative knowledge acquired by the students of CG and EG after attending the first three lectures. Data analysis was based on the operational measure termed "coefficient of proficiency" (Bespalko, 1989). An example of the measure for an individual student is provided here. For the part gauging declarative knowledge, student #1 from CG had 8 correct answers out of 12, i.e. giving a measure for the coefficient of proficiency of 0.67 (8/12). The similarity of the two groups was assessed using summary statistics. The coefficient of proficiency in the two groups before the experiment is presented in Table 1.

Table 1: Summary statistics of coefficient of proficiency in the two groups before the experiment

	Control Group (CG)	**Experimental Group (EG)**
N	17	17
Mean	0.58	0.57
Standard deviation (SD)	0.144	0.148
Median	0.58	0.58
Min – Max	0.42 – 0.83	0.42 – 0.83

From Table 1, the mean values of coefficient of proficiency were 0.58 (range: 0.42 – 0.83) in CG, and 0.57 (range: 0.42 – 0.83) in EG. Therefore, the groups were deemed comparable and ready for the empirical training.

The first question we examined after the empirical training was the possible relationship between formative assessment and students' acquisition of lecture material. In other words, we wanted to get evidence of whether there is a within-group change in the coefficient of proficiency in the post-training assessment compared to the pre-experimental assessment. The following null hypothesis was developed.

H01. There is no within-group change in the coefficient of proficiency in the post-training assessment compared to the pre-experimental assessment.

To test the first null hypothesis, we compared the coefficient of proficiency in each group before and after experimental interventions (different training procedures), and then used a statistical test of significance (two-sample t-test) to assess the group difference. Table 2 below shows individual data, as well as group summary statistics (mean, SD, median, minimum, and maximum) before the experimental intervention (Pre), after the experimental intervention (Post), and the change (Post minus Pre).

Table 2: Individual data and group summary statistics of coefficient of proficiency before the experimental intervention (Pre), after the experimental intervention (Post), and the change (Post minus Pre)

	Control Group (CG)				**Experimental Group (EG)**		
Student	**Pre**	**Post**	**Change**	**Student**	**Pre**	**Post**	**Change**
1	0.67	0.63	-0.04	1	0.67	0.83	0.16
2	0.42	0.5	0.08	2	0.42	0.67	0.25
3	0.58	0.59	0.01	3	0.58	0.83	0.25
4	0.5	0.67	0.17	4	0.83	0.92	0.09
5	0.75	0.79	0.04	5	0.75	0.96	0.21
6	0.42	0.5	0.08	6	0.42	0.63	0.21
7	0.58	0.59	0.01	7	0.58	0.75	0.17
8	0.75	0.67	-0.08	8	0.42	0.67	0.25
9	0.42	0.5	0.08	9	0.83	0.87	0.04
10	0.83	0.84	0.01	10	0.5	0.67	0.17
11	0.75	0.75	0	11	0.42	0.5	0.08
12	0.42	0.5	0.08	12	0.42	0.63	0.21
13	0.67	0.75	0.08	13	0.67	0.83	0.16
14	0.42	0.5	0.08	14	0.5	0.71	0.21
15	0.58	0.67	0.09	15	0.42	0.63	0.21
16	0.67	0.79	0.12	16	0.67	0.92	0.25
17	0.42	0.59	0.17	17	0.67	0.75	0.08
Mean	0.58	0.64	0.06	**Mean**	0.57	0.75	0.18
SD	0.144	0.117	0.067	**SD**	0.148	0.128	0.067
Median	0.58	0.63	0.08	**Median**	0.58	0.75	0.21
Minimum	0.42	0.50	-0.08	**Minimum**	0.42	0.5	0.04
Maximum	0.83	0.84	0.17	**Maximum**	0.83	0.96	0.25

In CG, the mean change in the coefficient of proficiency was 0.06 (range: -0.08 to 0.17). That is, on average, the coefficient of proficiency in CG increased by 0.06 points, whereby there was a student (#8 in CG) whose change in the coefficient of proficiency was -0.08 (decrease of 0.08 points) and the maximum observed increase was 0.17 points (students #4 and #17 in CG). In EG, the mean change in the coefficient of proficiency was 0.18 (range: 0.04 to 0.25). In essence, on average, the coefficient of proficiency in EG increased by 0.18 points, with all students in this group exhibiting positive changes, from the minimum of 0.04 points (student #9 in EG) to the maximum of 0.25 points (students #2, #3, #8, and #16 in EG). To test the significance of within-group changes in coefficient of proficiency, a paired t-test was applied on the individual differences (post- minus pre-), for each treatment group. The results are displayed in Table 3,

columns 3 – 5. The mean changes in both groups were significantly different from zero. In CG, the mean change (95% confidence interval) was 0.06 (0.02 - 0.09), p = 0.0027, and the corresponding values for EG were 0.18 (0.14 - 0.21), p < 0.0001. Hence, the null hypothesis H01 was rejected, and the conclusion was drawn that there is a within-group change in the coefficient of proficiency in the post-training assessment compared to the pre-experimental assessment, in both CG and EG.

The second question addressed was the possible relationship between measuring procedures' content and students' performance, i.e. we wanted to figure out whether there was a significant group difference in the change of coefficient of proficiency at the post-training assessment. The following hypothesis was developed.

H02. There is no between-group difference in the change of coefficient of proficiency at the post-training assessment.

To test the second null hypothesis, a two-sample t-test was performed on individual values of change in the coefficient of proficiency from pre- to post-. Figure 2 displays these individual changes by treatment groups. There is evidence of a greater increase in the coefficient of proficiency in EG compared to CG. Table 3 (columns 6 – 8) shows the results of the two-sample t-test. The group mean difference (EG vs. CG) was 0.12 points, 95% confidence interval (0.07 - 0.17), two-sample t-test p < 0.0001. Therefore, EG resulted in a mean difference of 0.12 points compared to CG, a statistically significant and meaningful increase.

Table 3: Results of statistical tests to evaluate the significance of changes (Post- minus Pre-) within each group, and the group difference (EG vs. CG)

		Change (Post- minus Pre-)			Group difference (EG vs. CG)		
Group	**N**	**Mean**	**95% CI**	**p-value**	**Mean**	**95% CI**	**p-value**
Control (CG)	17	0.06	(0.02 - 0.09)	0.0027[a]	0.12	(0.07 - 0.17)	<0.0001[c]
Experimental (EG)	17	0.18	(0.14 - 0.21)	<0.0001[b]			

[a] paired t-test, $t = 3.5406$, $df = 16$
[b] paired t-test, $t = 10.802$, $df = 16$
[c] two-sample t-test, $t = 5.1518$, $df = 32$

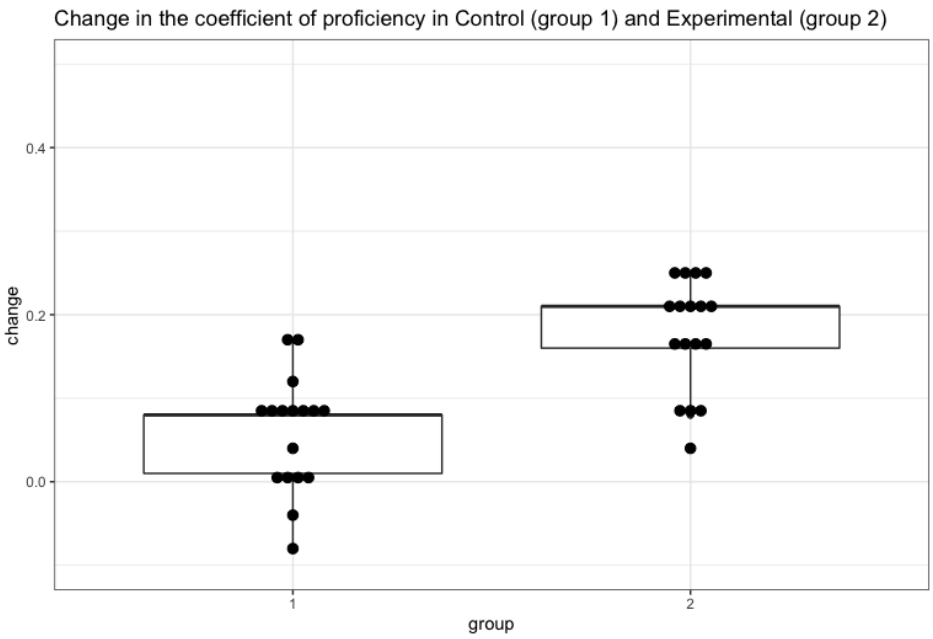

Figure 2: The evidence of a greater increase of the coefficient of proficiency in EG in comparison with CG

The analysis of students' self-assessment surveys corroborated the positive impact of the empirical questionnaires in both groups. The results of students' grading of their readiness to deal with different types of questions are shown in Table 4.

Table 4: Students' self-perceived readiness to answer different types of questions

I am ready	CG (number of students)					EG (number of students)				
	always	very often	some-times	rarely	never	always	very often	some-times	rare ly	never
to state facts	1	7	9	0	0	5	8	4	0	0
use definitions in deciding the choice	1	7	9	0	0	5	8	4	0	0
to classify sentences into true or false	1	7	9	0	0	5	8	4	0	0
to answer questions in written form	1	7	9	0	0	5	8	4	0	0
to describe the steps to be taken to solve a problem	1	7	9	0	0	5	8	4	0	0
to combine rules to propose an approach to problem solving	1	7	9	0	0	5	8	4	0	0
to express my point of view orally	1	7	9	0	0	5	8	4	0	0

When asked about the effectiveness of empirical questionnaires 100% of the students from CG and EG answered that the tasks encouraged their thinking, reasoning and awareness, helped to understand and remember lecture material. However, we observed some differences between the groups. The number of students who answered "always" or "very often" was 8 out of 17 (47%) in CG, and it was 13 out of 17 (76%) in EG (Fisher's exact test for group difference 2-sided $p = 0.1571$). These results speak in favour of EG, supporting the results from Table 3.

During the discussion of the final test results, 14 students out of 34 (10 out of 17 (59%) from CG, and 4 out of 17 (24%) from EG) stated that they had difficulty in performing the tasks related to the steps of problem solving (simple rule questions) and combining rules (high order rules tasks). The students explained that the tasks were new and puzzling for them. Though they knew the theory, they doubted how to design educationally oriented recommendations.

Discussion

The aim of this research was to further study the relationship between formative assessment and learners' cognitive skills. A review of the literature identified a clear link between sound assessment and student achievement. It was thought that further study of the relationship between assessment procedures and student performance would provide additional data related to understanding more clearly the topic of cognitive skills assessment for prospective language teachers. Our findings correspond to Brown, Bull and Pendlebury's idea (1997) that if you want to change student learning then change the methods of assessment. The study focused on three major research questions. What procedures could be used to assess students' learning outcomes from a cognitive perspective? Does a relationship exist between formative assessment and students' acquisition of lecture material? Is there a relationship between the measuring procedures' content and students' performance?

Thirty-four third-year students were identified to participate in research teaching, which lasted ten weeks (one lecture per week). The students were randomly divided into CG and EG. After each lecture, for self-control, both groups got empirical questionnaires, which differed in content.

Answering Question 1, we present evidence that evaluation of learning outcomes could be embedded within a theory of cognitive skills development provided by Anderson (1982) and a taxonomy of important dimensions for assessing achievement, proposed by Glaser, Lesgold and Lajoie (1987). We showed that one of the indexes of cognitive skills development was knowledge organisation and structure, and that the procedures for assessing cognitive skills could be organised around a framework of declarative and procedural knowledge. This agrees with the results of previous research, which argue the effectiveness of such an organisation. The proposed model has survived empirical probing (Denysiuk & Stokaz, 2018; Elton & Johnston, 2002), it has guided the development of instructional activities (Ismail, 2016; Kaba & Ramaiah, 2020), and it has been widely recognised as one of the most important theoretical developments of the cognitive revolution (Royer, Cisero & Carlo, 1993; Li, Hunter & Lei, 2016).

Question 2. The obtained results confirmed the interrelationship between formative assessment and students' acquisition of lecture material. The study found that after the work with empirical questionnaires and participation in formative assessment, students scored better when compared to a pre-research period.

The results, presented in Table 2, demonstrate that experimental training positively influenced knowledge acquisition both in CG and EG. Findings support previous research works, which show that formative assessment allows students to identify their strengths and weakness throughout studies and monitor their progress towards achieving learning objectives (O'Keeffe et al., 2020; Vionea, 2018; Wang et al., 2018).

Question 3. The results, presented in Table 3, showed a significant difference between CG and EG (the group mean difference in EG vs. CG was 0.12 points). Pedagogical observation revealed that the procedures, applied in EG more helped students to be able to remember facts (multiple-choice tests); be sensitive in domain knowledge (Sentence Verification Technique tests); put their feedback into words without restricting their thoughts(open-ended questions); and, according to Jackson et al., (2018), leverage natural-language processing to provide adaptive follow-up prompts that target particular information (conversation-based questions). This correlates with Regian and Schneider's (1990), idea that assessment activities targeted at task-specific cognitive process are much better than traditional procedures. The results of this study when viewed overall could suggest that traditional procedures may be effective. However, noting the impact of cognitive assessment procedures, we argued that even though both CG and EG made improvements, the growth in test scores was significantly higher for EG (Table 3).

Although this study is limited to the course Language Teaching Methodology, and generalisations cannot be made to all courses, the findings suggest that some relationship exists between cognitive formative assessment techniques and improved student performance on the final test. At the same time, it should be noted that there are some concerns that accompany the use of cognitive skills measure procedures as a means of assessing training success. According to Royer, Cisero and Carlo (1993), "inferences about individual accomplishment should only be made based on measurement procedures that are highly reliable and that have accumulated a mosaic of evidence consistent with the interpretation that the measure is valid for a specific purpose" (236).

Implications. The implications of this study are tentative in nature and relate to the following issues. 1. Final grades should be based on more than one assessment. Development throughout the course appears to be effective, and it results in improved student scores on the final exam. 2. It is plausible that a student's "success rate" at the finals based not only on a pure theoretical background, but measured in terms of cognitive competency may be taken into account by school administration trying to hire the best specialists. 3. The implications seem to relate to specific courses and academic curriculum, which have a profound influence on the methodological approaches to teaching in schools.

Limitations. There are some important limitations of the study. Firstly, the study sample size is fairly small. Secondly, threats to internal validity may have occurred because of experimental effects. The students in CG and EG knew that they were in an experiment; they might have adapted their behaviour in a way that prevented unbiased estimation of the treatment effect. Thirdly, the instrument with its classifications into Gagne's learning hierarchy is an imprecise and insufficiently specific measurement device. It is always disputable whether the question/task totally corresponds to a certain level of hierarchy. Instrument classification errors may have had an impact on the results. Fourthly, the research requires a longitudinal study.

Conclusions.

The research reviewed in this article indicates that the suggested set of measurement procedures can provide a reliable means of assessing learning outcomes. Specifically, the procedures are sensitive to instruction, which aims at boosting career-focused educational environment; they can be used as a means of formative assessment of learners' cognitive skills; as a tool of predicting future student academic performance; and as an instrument for assisting curriculum placement decisions.

The research reviewed in the article certainly does not exhaust the realm of issues involving cognitive skills testing. There is a variety of questions that could be asked, among which is what procedures can be used to measure other Glaser's dimensions, such as depth of problem representation; quality of mental models; efficiency; automaticity; metacognitive skills for learning. Future research on these issues could contribute to expanded uses for cognitive skills testing.

References:

Ainsworth, J. (2013). Sociology of Education: an A-to-Z Guide. Two Volume Set. Georgia State University, USA, SAGE Publications, Inc Encyclopedia Definition of cognitive skill. Retrieved 4 May 2021 from https://www.amazon.com/Sociology-Education-Z- Guide/dp/1452205051.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89(4),* 369-406. https://doi.org/10.1037/0033-295X.89.4.369

Benton, St. L., & Kiewra, K. A. (1987). The Assessment of Cognitive Factors in Academic Abilities. In Royce R. Ronning, John A. Glover, Jane C. Conoley, University of Nebraska-Lincoln; Joseph C. Witt Louisiana State University (Eds.), *The Influence*

of Cognitive Psychology on Testing (pp. 148-200). Lawrence Erlbaum Associates, Publishers: Hillside, New Jersey London. Retrieved 19 May 2021 from https://digitalcommons.unl.edu/buroscogpsych/9/

Bespalko, V. P. (1989). *Slagaemye pedagogicheskoj tehnologii* [Components of pedagogical technology], Moscow: Pedagogika.

Bonner, S., & Pennington, N.(1991). Cognitive Processes and Knowledge as Determinants of Auditor Expertise. *Journal of Accounting Literature, 10,* 1-50.

Boruff, J. T., & Harrison, P. (2018). Assessment of knowledge and skills in information literacy instruction for rehabilitation sciences students: a scoping review. *Journal of the Medical Library Association, 106(1),* 15-37. https://doi:10.5195/JMLA.2018.227

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education.* London: Routledge.

Cole, M. W. (2010). *Influence of Assessment for Learning Professional Development in Rural Georgia Public Schools.* A Dissertation Presented to The Faculty of the School of Education Liberty University. Retrieved 19 May 2021 from https://www.routledge.com/Assessing-Student-Learning-in-Higher-Education/Brown-Bull-Pendlebury/p/book/9780415144605

Denysiuk, O. V., & Stokaz, Ya. M. (2018). Metodychnyy pidkhid do formuvannya ta rozvytku stratehichnykh znan′ pidpryyemstva [A Methodological Approach to Forming and Developing Strategic Knowledge of an Enterprise]. *The Problems of Economy, 2,*191-198. https://www.problecon.com/

Elton, L., & Johnston, B. (2002). *Assessment in universities: a critical review of research.* UK: Learning and Teaching Support Network Generic Centre. Retrieved 19 May 2021 from https://eprints.soton.ac.uk/59244/

Gagne, R. M., & Medsker, K. L. (1996). *The Conditions of Learning: Training Applications.* Fort Worth: TX: Harcourt Brace College Publishers.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a Cognitive Theory for the Measurement of Achievement. In Royce R. Ronning, John A. Glover, Jane C. Conoley, University of Nebraska-Lincoln; Joseph C. Witt Louisiana State University (Eds.), *The Influence of Cognitive Psychology on Testing* (pp. 41-85). Lawrence Erlbaum Associates, Publishers: Hillside, New Jersey London.

Ismail, Y. (2016). The use of problem solving techniques to develop semiotic declarative knowledge models about magnetism and their role in learning for prospective science teachers. *International Journal of Educational Administration and Policy Studies,* 5(8), 97-109. https://doi.org/10.5897/ijeaps2016.0476

Jackson, T. G., Castellano, K. E., Brockway, D., & Lehman, B. (2018). Improving the Measurement of Cognitive Skills Through Automated Conversations. *Journal of Research on Technology in Education, 50(3),* 226-240. https://doi.org/10.1080/15391523.2018.1452655

Jönsson, A. (2020). Definitions of Formative Assessment Need to Make a Distinction Between a Psychometric Understanding of Assessment and "Evaluative Judgment". *Frontiers in Education,* 5(2). https://doi: 10.3389/feduc.2020.00002.

Kaba, A., & Ramaiah, K. Ch. (2020). Measuring Knowledge Acquisition and Knowledge Creation: A Review of the Literature. *Library Philosophy and Practice (e-journal).* Retrieved 4 May 2021 from https://digitalcommons.unl.edu/libphilprac/4723/

Lenz, K., Dreher, A., Holzapfel, L., & Wittmann, G. (2020). Are conceptual knowledge and procedural knowledge empirically separable? The case of fraction. *British Journal of Educational Psychology, 90(3),* 809-829. https://doi.org/10.1111/bjep.12333

Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing, 33,* 391-409. https://doi.org/10.1177/0265532215590848

McIlwain, D. J. F., & Sutton, J. (2015). Methods for measuring breadth and depth of knowledge. In J. Baker, & D. Farrow (Eds.), *Routledge handbook of sport expertise* (pp. 221-231). London; New York: Routledge, Taylor and Francis Group. https://doi.org/10.4324/9781315776675-20

O'Keeffe, L., Rosa, A., Vannini, I., & White, B. (2020). Promote Informal Formative Assessment practices in Higher Education: the potential of video analysis as a training tool. *Form@re Open Journal Per La Formazione in Rete,* 20(1), 43-61. https://doi.org/10.13128/form-8241

Panadero, E., Jonsson, A. & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four metaanalyses. *Educational Research Review, 22,* 74. https://doi.Org/10.1016/j.edurev.2017.08.004

Piaget, J. (1977). *The Development of Thought: Equilibration of Cognitive Structures.* UK, Oxford: Blackwell.

Regian, J. W., & Schneider, W. (1990). Assessment procedures for predicting and optimizing skill acquisition after extensive practice. In N. Frederiksen, R. Glaser, A. Lesgold & M.G. Shafto (Eds.), *Diagnostic monitoring of skill acquisition and knowledge acquisition* (pp. 297-323). Hillsdale, NJ: Erlbaum. https://doi.org/10.4324/9780203056899-16

Richter-Beuschel, L., Grass, I., & Bogeholz, S. (2018). How to Measure Procedural Knowledge for Solving Biodiversity and Climate Change Challenges. *Education Sciences, 8,* 190. https://doi.org/10.3390/educsci8040190.

Royer, J. M., Cisero, Ch. A., & Carlo, M. S. (1993). Techniques and Procedures for Assessing Cognitive Skills. *Review of Educational Research, 63(2),* 201-243. https://doi.org/10.3102/00346543063002201

Tigelaar, D., & Sins, P. (2020). Effects of formative assessment programmes on teachers' knowledge about supporting students' reflection *Journal of Vocational Education & Training,* 1-24. https://doi.org/10.1080/13636820.2020.1726992

Ullah, Z., Lajis, A., Jamjoom, M., Altalhi, A. H., Shah, J., & Saleem, F. (2019) A Rule-Based Method for Cognitive Competency Assessment in Computer Programming Using Bloom's Taxonomy. In *IEEE Access, 7,* 64663-64675. https://doi.org/10.1109/ACCESS.2019.2916979

Vandierendonck, A (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods, 49,* 653-673. https://doi.org/10.3758/s13428-016-0721-5

Voinea, L. (2018). Formative Assessment as Assessment for Learning Development. *Revista de Pedagogie - Journal of Pedagogy, I.XVI(\),* 7-23. https://doi.org/10.26755/revped/2018.1/7

Wang, D., Sun, Y., & Jiang T. (2018). The Assessment of Higher Education Quality from the Perspective of Students through a Case Study Analysis. *Frontiers of Education in China, 13(2),* 267-287. https://doi.org/10.1007/s11516-018-0014-0.

Vygotsky, L.S. (1962). *Thought and Language.* MIT Press, Cambridge, Mass, USA.

Vygotsky, L.S. (1978). *Mind in Society: The Development of Higher Psychological Process.* Harvard University Press, Cambridge, Mass, USA.

Received: May 6, 2021
Accepted: June 12, 2021