Computers in Human Behavior 75 (2017) 81–91

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Promoting collaborative learning through regulation of guessing in clickers

I.M. Kulikovskikh [a, *], S.A. Prokhorov [a], S.A. Suchkova [b]

[a] Information Systems and Technologies Department, Samara National Research University, 34 Moskovskoe sh., 443086 Samara, Russia
[b] Foreign Languages Department, Samara National Research University, 34 Moskovskoe sh., 443086 Samara, Russia

A R T I C L E I N F O

Article history:
Received 23 May 2016
Received in revised form
27 March 2017
Accepted 1 May 2017
Available online 8 May 2017

Keywords:
Partial guessing
Regulation
Collaborative learning
Clickers

A B S T R A C T

Collaborative learning is a promising avenue in education research. Learning from others and with others can foster deeper learning at a multiple-choice assignment, but it is hard to control the level of students' pure guessing. This paper addresses the problem of promoting collaborative learning through regulation of guessing when students use clickers to answer multiple-choice questions of various levels of difficulty. The study is aimed at identifying how the difficulty of the task and students' levels of knowledge influence on the degree of partial guessing. To answer this research question, we developed two research models and validated them by testing 84 students with regard to the students' level of knowledge and the penalty announcement. The findings of this research reveal that: a) the announcement of penalty has a negative effect on promoting collaborative learning even if it leads to reducing pure guesses in test results; b) questions that require higher-order thinking skills promote collaborative learning to a greater extent; c) creating mixed level groups of students seems advisable to enhance learning from collaboration and, thus, to decrease the degree of pure guessing.

1. Introduction

Collaborative learning is a pedagogical approach, which helps to enhance learning performance (Blasco-Arcas, Buil, Hernández-Ortega, & Javier Sese, 2013; McDonough & Foote, 2015). In-depth research indicates that this type of learning environment leads to deeper learning while students teach each other by addressing misunderstanding and clarifying misconceptions. In the collaborative learning environment, students gain different perspectives and, thus, articulate and defend their own ideas. It was Lev Vygotsky who laid the foundations for collaborative learning (Vygotsky, 1978). His concept of learning, called the zone of proximal development, cast doubt on knowledge-based tests as a proper means to measure students' level of knowledge. Vygotsky contended that, in order to gauge the level of true knowledge, it is required to examine an ability to solve problems both independently and in a group. But measuring the knowledge of students who are working in a group is a complicated problem.

One way of stimulating peer collaboration and, at the same time, measuring individual performance is using clickers (Brady, Seli, & Rosenthal, 2013; Chien, Chang, & Chang, 2016; Cook & Calkins, 2013; Lantz & Stawiski, 2014; Lantz, 2010; Mayer et al., 2009). Some studies highlight the effectiveness of this method because it promotes active learning through student engagement (McDonough & Foote, 2015). For instance, in their research, Smith et al. (2009) used clickers to test in-class concept questions. At first, students were asked to answer a question after a peer discussion. Then, they were posed a similar clicker question, but they followed the instruction to give an answer independently. Smith et al. (2009) analyzed the improved percentage of correct answers after peer discussion. The authors offered two possible explanations for higher grades: the result of conceptual understanding or simply the outcome of choosing the answer most supported by more knowledgeable peers. The authors concluded that the peer discussion led to better understanding even when none of the students knew the correct answer. Although this research shed light upon the major problem of distinguishing between actual learning from students' collaboration and the influence of more prepared students on their peers, there seemed to be some problems with assessing learning performance accurately.

This assessment problem partly results from the limitations imposed by the testing format: clickers are traditionally used in multiple-choice testing (Little & Bjork, 2016). There are two major

* Corresponding author.
E-mail address: kulikovskikh.i@gmail.com (I.M. Kulikovskikh).

http://dx.doi.org/10.1016/j.chb.2017.05.001
0747-5632/

issues of multiple-choice testing, which are widely debated in the literature.

The first issue is designing questions which go beyond Bloom's lower-order thinking levels: *recalling*, *understanding*, and *applying* to the higher-order levels: *analyzing*, *evaluating*, and *creating* (Anderson et al., 2001; Bloom et al., 1956). On the one hand, some studies (Anderson et al., 2001; Mayer, 2002; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2010; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2011; Thelwall, 2000) point out that it is possible to design multiple-choice quizzes that test higher-order thinking skills. On the other hand, some researchers argue that multiple-choice assignments are deemed to measure only factual recalling (Butler & Roediger, 2008; Nickerson, Butler, & Carlin, 2015; Nicol, 2007). Therefore, many instructors offer the easiest way to manipulate test difficulty, i.e. to vary the number of multiple-choice alternatives (Butler & Roediger, 2008; Dehnad, Nasser, & Hosseini, 2014; Lesage, Valcke, & Sabbe, 2013; Tarrant & Ware, 2010). But an increase in the number of distractors may lead to a decrease in proportions of correct responses. Students are likely to acquire false knowledge instead of enhancing retention of the material. As a result, such test format may increase students' exposure to misinformation. Butler & Roediger (2008) indicate that a distractor has the most detrimental effect unless proper feedback is provided Nicol (2007).

An opposing view is suggested in Bjork's recent research (Bjork, Little, & Storm, 2014; Bjork, Soderstrom, & Little, 2015; Little & Bjork, 2015), where it is stated that multiple-choice testing can promote deep learning and increase long-term retention even when no corrective feedback is given. In accordance with these studies, multiple-choice testing can stimulate the type of retrieval processes known to improve learning (Bjork et al., 2015). In this case, instructors need to provide students with a metacognitive strategy to encourage more complex thinking. This strategy is aimed at considering all the alternatives to cogitate not only why the selected answer is correct, but also why distractors are incorrect. Moreover, students should engage in this metacognitive strategy even if they are certain what answer is correct.

However, applying metacognitive strategies may pose the other serious assessment problem: if students can eliminate some responses based on critical analysis, they can get the correct answer with partial guessing, the level of which is often difficult to assess correctly (Ben-Simon, Budescu, & Nevo, 1997; Kubinger, Holocher-Ertl, Reif, Hohensinn, & Frebort, 2010). An extensive body of literature puts forward different scoring procedures to examine partial guessing (Arnold & Arnold, 1970; Bereby-Meyer, Meyer, & Budescu, 2003; Espinosa & Gardeazabal, 2010; Lord, 1980). The primary purpose of these methods is to alleviate pure guessing effects on multiple-choice items and, thus, to reveal students' true knowledge. For instance, Ghafournia (2013) attempted to approach this problem analysing test-taking strategies in answering multiple-choice tests at three levels of English proficiency. The author studied the following subcategories of strategies: time management, error avoidance, guessing, and intent consideration (Ghafournia, 2013). The findings of this research demonstrate significant differences only in using guessing strategies across the three levels of proficiency. While the higher level students used the error avoidance strategy and the time management strategy more frequently, the lower level students employed the guessing strategy less regularly. In contrast to the results of the lower level group and the higher level group, the intermediate level students used the guessing strategy to a much greater extent. These results could be interpreted as follows. The higher level students have a sufficient level of knowledge to answer questions, so they do not need to heavily rely on the guessing strategy. By contrast, the lower level students take pure guesses as they may not have enough knowledge to adopt guessing as a strategy. Finally, the intermediate level students have only partial knowledge. As a result, they demonstrate some partial guessing in attempt to avoid distractors. Consequently, the level of guessing depends not only on the order of thinking skills, but also on the level of students' knowledge.

What is not specifically tackled in the studies reviewed above is how the levels of cognition and students' levels of knowledge influence on the degree of guessing. This is the research question raised in this study. Addressing this gap with regard to collaborative learning, we stated the **objective** to look into the problem of promoting collaborative learning through regulation of guessing in answering clicker questions. Firstly, we support the idea that clickers can be seen as an effective instrument for promoting deeper understanding and improved students' performance via collaboration. Clickers can help to develop students' critical thinking skills (Blasco-Arcas et al., 2013; Brady et al., 2013; Levesque, 2011), especially when designed questions are based on a taxonomy to encourage higher-order thinking (Bode, Drane, Ben-David Kolikant, & Schuller, 2009; Bruff, 2009; Cook & Calkins, 2013). Secondly, the process of collaboration is not limited to applying only cognitive and metacognitive strategies. It also involves such aspects as social and metasocial interaction (Wang, Wallace, & Wang, 2017). Consequently, the regulation of this process is crucial for creating an effective learning environment. Though there is research into different types of regulation (De Backer, Van Keer, Moerkerke, & Valcke, 2016; Jarvela & Hadwin, 2013; Jarvela & Hadwin, 2015; Jarvela, Malmberg, & Koivuniemi, 2016; Raes, Schellens, De Wever, & Benoit, 2016; van Leeuwen, Janssen, Erkens, & Brekelmans, 2015; Winne, 2015), which is primarily focused on developing skills of self-regulation (Grau & Whitebread, 2012), co-regulation (Chan, 2012) and socially shared regulation (De Backer, Van Keer, & Valcke, 2014; Isohatala, Jarvenoja, & Jarvela, 2017; Jarvela & Hadwin, 2015; Malmberg, Jarvela, Jarvenoja, & Panadero, 2015), but it seems little attention is paid to the problem of *guessing* regulation.

To achieve our research aim, we tested two control groups of students: lower level students and higher level students. They were given a set of clicker questions, increasing in difficulty and involving both lower-order thinking skills (LOTS) and higher-order thinking skills (HOTS). During the tests, all the students were encouraged to collaborate. However, some of them were announced the penalty for guessing, while the others had no penalty. In addition, the students were given bonus points for answering the clicker questions correctly, so they had an incentive to take the questions seriously.

This paper is organised as follows. Section 2 presents the hypotheses and the proposed research models. It also describes the tests, participants, and procedure used to support the present research. Section 3 reveals the results of descriptive statistics in support of the hypotheses. Section 4 summarises the findings of this study and answers the raised research question.

2. Method and materials

This section discusses the hypotheses formulated to examine the research question and the research models created to visually represent the logic behind the hypotheses. Then, we provide a description of tests, participants, and procedure used to support the present research.

2.1. Hypotheses

To answer the research question, stated in Section 1, we first consider the relationship between the order of thinking skills and the degree of guessing. We hypothesize:

H1. The lower the order of thinking skills is, the more partial guesses in test results are.

H2. The higher the order of thinking skills is, the more pure guesses in test results are.

We introduce penalty as a method for regulating guessing in answering the clicker questions. To address the proposed hypotheses with regard to the announcement of penalty for guessing, we assume that students, working in groups, may be open to collaboration to a different extent. Thus, we also add two more hypotheses for examination:

H3. Testing questions that require lower-order thinking skills promotes learning from collaboration to a lesser extent in both cases: with penalty and no penalty for guessing.

H4. Testing questions that require higher-order thinking skills promotes learning from collaboration to a greater extent if there is no penalty for guessing.

Finally, we consider the research question from the viewpoint of the levels of students tested.

H5. Lower level students make more pure guesses and learn from collaboration to a lesser extent.

H6. Higher level students make more partial guesses and learn from collaboration to a greater extent.

2.2. Research models

To explain the logic behind the formulated hypotheses (see Section 2.1), we developed research models depicted in Fig. 1. The model shown in Fig. 1a describes the behavior of lower level students in answering clicker questions. In turn, Fig. 1b models the behavior of higher level students. As can be seen, the research models connect a thinking skills line and a guessing line. The thinking skills line ranges from LOTS to HOTS, while the guessing line measures the degree of guessing between *pure guesses* and *no guesses* with *partial guesses* in borderline cases.

In the present study, we took two arbitrary points on the thinking skills line: the first point illustrates the combination of the questions that require LOTS; the second point integrates all the questions that require HOTS. A solid arrow, which has the endpoint at one of these arbitrary points, signifies some value on the guessing line when no penalty was announced. A dashed arrow, pointing to the guessing line, corresponds to the case of announced penalty. To illustrate the impact of collaboration on the guessing

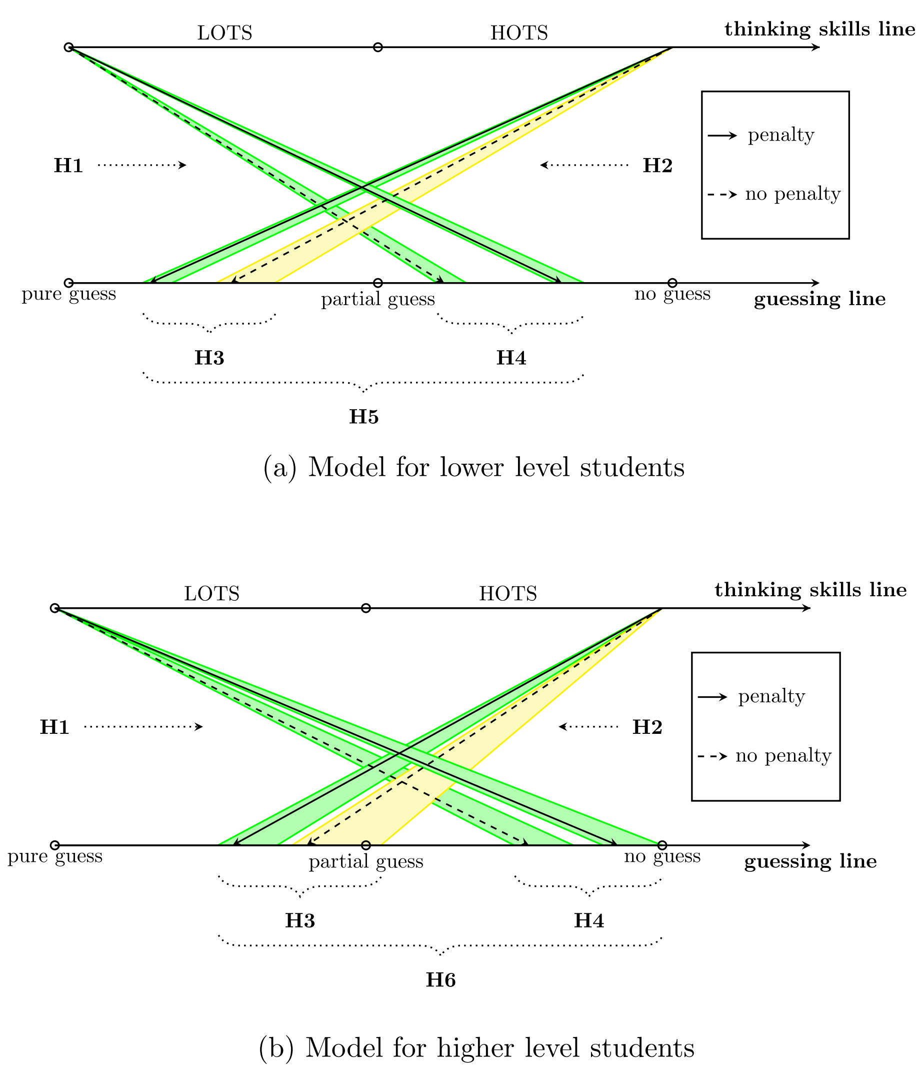

Fig. 1. The proposed research models based on the hypotheses H1–H6.

line, we used the idea of "footprint of uncertainty" borrowed from 2-type fuzzy logic sets (Bustince, Fernandez, Hagras, Pagola, & Barrenechea, 2014). Instead of taking a single value on the guessing line, which indicates the result of guessing without collaboration, we use a range of values to demonstrate the influence of collaborative learning. Consequently, the stronger the impact of collaboration is, the greater the width of "footprint" on the guessing line becomes. While all the "footprints" are colored in green, the "footprint" with the strongest impact is highlighted in yellow. It should be noted that the influence of collaboration is mostly positive, so the "footprints" are skewed right relative to the arrows.

Now, if we turn to the hypotheses, we can see that the slope of the arrows and their positions on the guessing line in Fig. 1 present the hypotheses H1 and H2. Then, the width of "footprints" on the guessing line illustrates the hypotheses H3 and H4. Finally, the positions of "footprints" and the total width of all the "footprints" visually present the hypotheses H5 and H6, which reflect an essential difference in the behavior of lower level students (see Fig. 1a) and the behavior of higher level students (see Fig. 1b).

2.3. Measures

2.3.1. Design

The clicker questions related to Bloom's taxonomy tiers (Anderson et al., 2001; Bloom et al., 1956) were designed according to the comprehensive and explicit recommendations (Anderson et al., 2001; Lord, 1980; Mayer, 2002; Thelwall, 2000) on how to assess the required level of thinking skills. Moreover, the levels *applying*, *analyzing*, and *evaluating*, which are often difficult to recognize accurately, were provided with additional instructions to help to produce better learning outcomes. The test instructions can be found in Appendix.

Each of the tests includes a number of items, given in brackets, which, in turn, consist of different number of distractors. In addition, these tests comprise both single-answer items and multiple-answer items. To invoke beneficial retrieval processes in students, the tests contain competitive incorrect distractors (Bjork et al., 2015).

2.3.2. Instructions

The main purpose of this testing is to examine partial guessing (or partial knowledge) (Aluisio, Aquino, Pizzirani, & de Oliveira, 2003; Zareva, 2012) assessing not only LOTS, but also HOTS. The topic of English prepositions seems a valid area to opt for. We now uncover the reasons for making this point.

Some of the meanings of basic prepositions are easy to demonstrate, and students can often guess their meanings. As a result, ESL/EFL learners often omit and misuse such prepositions that may result in altering the meaning of the sentences (Jimánez Catalán, 1996; Lo, Wang, & Yeh, 2004; Ngu & Rethinasamy, 2006). In attempt to explain this issue, Clark (1968) indicates that prepositions with overlapping contexts in sentences have overlapping free associations. Hence, English learners use their linguistic knowledge in composing sentences, giving free associations which stimulate cognitive processes.

Many researchers (Gucht, Willems, & Cuypere, 2007; Kleiner, 2005; Lindstromberg, 2010; Tyler & Evans, 2003) support the idea that the main problem of learning prepositions follows from polysemy. Tyler and Evans (2003) have tackled prepositional meanings with regard to the relationship between their spatial meaning and non-spatial extension. They focus on cognitively motivated prepositional meanings and different senses of a preposition associated with metaphorical extensions. While Tyler and Evans (2003) contend that "a good deal of information needed to establish normal interpretation of most sentences comes from cognitive processes, conceptual structure and background knowledge rather than the individual lexical items" (p. 16), Kleiner (2005) has cast doubt on this claim. Kleiner insisted on the importance of semantic compositionality in line with cognitive processing which remains absolutely essential in understanding utterance meaning. From this point of view, semantic knowledge can be considered as a set of individual lexical items in a language and the rules of their combination.

The cross-linguistic difference between these lexical packages and rules dwells on another source of difficulty in acquisition of English prepositions. It includes differences in the type of grammatical form and the way of segmenting spatial scenes (Lo et al., 2004; Tyler & Evans, 2003). As a consequence, this variation potentially affects the patterns of extension to non-spatial senses. Thus, it is not surprising that prepositions pose a challenge to ESL/EFL learners, but it is often equally hard for native speakers to articulate the reasons for selecting the appropriate preposition for a specific context. Mueller (2011) has explored the sensitivity of the frequencies of linguistic forms to test the claim that even advanced learners rely quite heavily on collocation knowledge in lieu of precise semantic knowledge. This study indicates that learners' performance can be measured regarding their knowledge of the preposition's meanings, ignorance of these meanings, and the knowledge of high-frequency collocations. Consequently, each learner demonstrates a unique set of competences that can be considered as partial knowledge.

2.3.3. Tests

Based on the comprehensive analysis of English prepositions, we designed the tests (see Appendix), which enabled us to assess:

- *Tests 1–5*: the knowledge of preposition meanings (Jimánez Catalán, 1996; Lo et al., 2004; Ngu & Rethinasamy, 2006);
- *Test 6*: the sensitivity to the frequencies of linguistic forms (Mueller, 2011);
- *Test 7*: the ability to relate the spatial and temporal meanings of English prepositions (Kemmerer, 2005);
- *Test 8*: the ability to sense the polysemy of the preposition OVER in spatial meanings (Kleiner, 2005; Tyler & Evans, 2003);
- *Test 9*: the ability to recognize the OVER polysemy in metaphorical extensions (Gucht et al., 2007; Lindstromberg, 2010).

2.4. Participants

We tested the proposed hypotheses using a sample of 84 university students. All the students major in engineering. They are under 30 years of age. These students voluntarily signed up to take the tests as an option to get a bonus to the course credit.

Although the research results provided in Smith et al. (2009) revealed that students' collaboration leads to better understanding of in-class concepts even when none of the students knows the correct answer, there still is a possibility that the answer might be chosen under the influence of more knowledgeable students. To lessen this chance, we divided all the participants into two control groups based on their CEFR (The Common European Framework of Reference for Languages) level ("International language standards — Cambridge English, 2016"): (1) those whose level is higher than B2; we will refer to them as "hl-" (higher level students); and (2) those whose level is lower than B2; we will refer to them as "ll-" (lower level students). Table 1 describes the study cohort: sex, age, and the level of knowledge.

In addition, each of the groups was randomly separated into two subgroups to provide them with an extra instruction: penalty for guessing ("-p", $p = 0.25$) was imposed on "hl-p" and "ll-p", while the others ("hl-np" and "ll-np") had no penalty ("-np", $p = 0$). The

Table 1
The description of the study cohort (K = 84).

Control groups	"hl-" group	"ll-" group
Gender [K (%)]		
female	13 (15.48)	11 (13.09)
male	23 (27.38)	37 (44.05)
Age [Mean (Std.dev.)]	22.24 (3.33)	19.48 (1.73)

penalty was imposed intentionally to give an external motivation tool.

2.5. *Methodology*

2.5.1. *Data collection*

The two groups of students ("hl-" and "ll-") were given 30 min to complete a set of clicker questions. The questions were given from easier (LOTS) to more challenging (HOTS). During the testing, all the students were encouraged to tackle the tasks collaboratively. In addition, extra points were given for correct answers, so that the students were motivated to succeed.

As indicated in recent research, there are no marked differences in test scores if students respond to polling questions, using either clickers or mobile devices (Stowell, 2015), mobile phones or computers (Huff, 2015). Thus, we asked the participants to use their mobile devices to complete the tests. They were instructed to install a mobile application, created by the authors, to join the polling as a guest by entering their name.

For clarity, the described procedure is visually presented in Fig. 2. We included the thinking skills line into this presentation to correlate it with the research models given in Section 2.2. But, instead of the guessing line, we introduced the borderline of students' level of knowledge to present the research question from the viewpoint of the tested groups.

We introduced two variables to define thinking levels. Section 2.2 describes the two arbitrary points, which identify the combinations of questions that require LOTS and HOTS, respectively. We intended to use the same names for these variables. Let LOTS denote the result of summing the responses to items 1–4 (see Appendix) divided by the number of alternatives and items. Then, the variable HOTS defines the result of summing the responses, but to items 5–9 over the number of alternatives and items.

2.5.2. *Data analysis*

An extensive body of literature (Barroso-Mendez, Galera-Casquet, Seitanidi, & Valero-Amaro, 2016; Dupuis, Khadeer, & Huang, 2017; Navimipour & Zareie, 2015; Valaei & Baroto, 2017; Willaby, Costa, Burns, MacCann, & Roberts, 2015; Winne, 2015) puts forward partial-least-squares structural equation modeling (PLS-SEM). This success may be attributed to the possibility to accurately assess direct and non-direct effects (latent variables) that may help to measure various psychological factors. In addition, PLS-SEM's statistical inference relies on bootstrapping (Streukens & Leroi-Werelds, 2016) that helps to work well with small samples.

Some studies (Rigdon, 2016; Ronkko, McIntosh, & Antonakis, 2015), however, indicate that the benefits of PLS have been greatly overestimated. Ronkko et al. (2015) have highlighted: 1) the misleading characterization of PLS as an SEM methods; 2) some limitations of PLS for global model testing; 3) misconceptions about the superior ability of PLS to deal with small sample sizes and non-normality and etc. As a result, this may cause the difficulty in interpreting loadings of latent variables.

Taking into account the mentioned drawbacks and the restricted sample of data (see Section 2.4), we adopted small inference procedures that provide reliable results under cross-sectional sampling (Basu, 2001): jackknife and boostrap methods. For small to moderate sample sizes the resampling estimates are significantly better than the asymptotic estimates: the resampling procedures guarantee the nominal confidence coefficient, but are narrower than asymptotic normal confidence intervals.

3. Results

This section demonstrates the results of descriptive statistical analysis in support of the hypotheses.

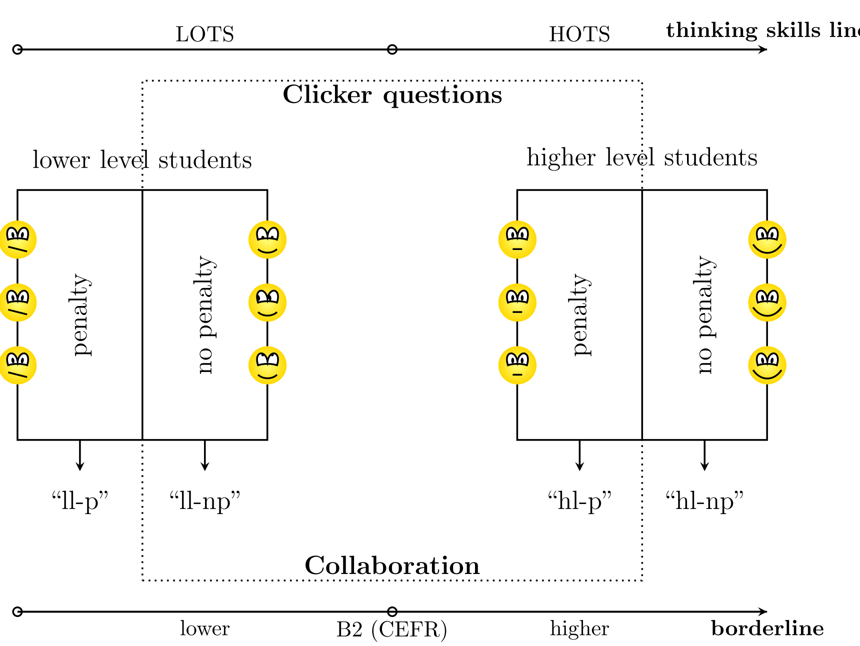

Fig. 2. The procedure of data collection.

Table 2
Means, standard deviations, *t*-test results, and Cronbach's α subject to the order of thinking skills.

Groups: hl-,ll- (-p,-np) (K = 84)					
	Descriptive statistics		*t*-test		Cronbach's α
	Mean	Std.dev.	t	*p*-value	
LOTS	0.49	0.11	9.19	0.0027^{**}	0.96
HOTS	0.25	0.07	7.31	0.0019^{**}	0.87

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

3.1. Tests reliability

The reliability and consistency of the test items were examined by Cronbach's α (see Table 2). As we can see, the values of Cronbach's α for both LOTS and HOTS variables (see Section 2.5.1) exceed 0.85. Consequently, the designed tests are sufficiently reliable for further detailed research.

Table 2 also presents the descriptive statistics (mean and std. deviation) and the *t*-test results. As expected, the mean of LOTS is higher than the mean of HOTS. On the other hand, the std. deviation of HOTS is lower than the same value for LOTS. Taking into account that the observed differences are statistically significant $p < 0.01$, this might be explained by both the influence of guessing and the result of students' collaboration. The students may have used guessing strategies to avoid some distractors when they answer the easier questions that require LOTS. On the other hand, when the students answer the more challenging questions that require HOTS, those who do not have enough knowledge may have taken part in collaboration to deepen their understanding. That may have led to reducing the std. deviation of HOTS. To distinguish between these possibilities, we analyzed the test results of different groups of students with regard to two conditions: the students' level of knowledge and the penalty announcement.

The results presented in Table 2 refer to all the participants ($K = 84$), where K is the number of students in a group. Based on the denotations proposed in Section 2.5.1, we defined the whole group of tested students as “hl + ll” or “p + np”.

3.2. Null hypothesis

To highlight a difference between the influence of guessing and the impact of collaborative learning, we first considered the descriptive statistics and the *t*-test results for each of the groups: “hl-p” ($K = 22$), “hl-np” ($K = 14$), “ll-p” ($K = 22$), “ll-np” ($K = 26$). The results of testing these groups are presented in Table 3 as well as are shown in Fig. 3a for LOTS and Fig. 3b for HOTS in the form of boxplots.

First, the presented results are also statistically significant: $p < 0.001$ (LOTS), $p < 0.01$ (HOTS) for “hl-” groups and $p < 0.05$ (LOTS, HOTS) for “ll-” groups. The lower p-value for “ll-” groups in testing both LOTS and HOTS questions signifies more pure guessing and less strong impact of collaboration. This can be attributed to the

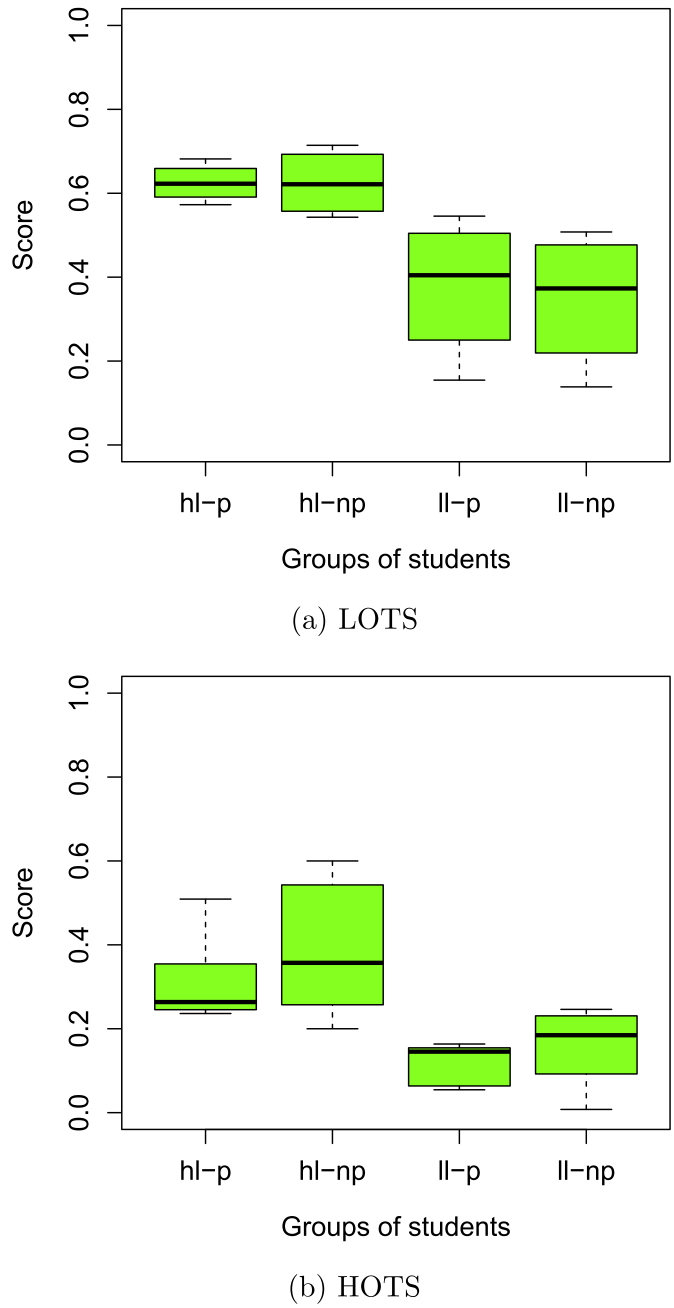

Fig. 3. Boxplots with regard to the order of thinking skills for each group: “hl-p”, “hl-np”, “ll-p”, “ll-np”.

Table 3
Means, standard deviations, and *t*-test results subject to the order of thinking skills for each group.

Group	Descriptive statistics				*t*-test			
	LOTS		HOTS		LOTS		HOTS	
	Mean	Std.dev.	Mean	Std.dev.	t	*p*-value	t	*p*-value
hl-p ($K = 22$)	0.63	0.05	0.32	0.12	27.19	0.0001^{***}	6.27	0.0033^{**}
hl-np ($K = 14$)	0.63	0.08	0.39	0.18	15.41	0.0006^{***}	5.01	0.0075^{**}
ll-p ($K = 22$)	0.38	0.17	0.12	0.05	4.45	0.0212^{*}	4.45	0.0211^{*}
ll-np ($K = 26$)	0.35	0.17	0.15	0.10	4.23	0.0242^{*}	4.23	0.0242^{*}

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

fact that “ll-” groups in collaboration, compared to “hl-” groups, produce less valuable and constructive ideas about what the correct answer is.

Let us take a closer look at the mean values and the std. deviation values for LOTS. As we can see, the mean of “hl-” groups are equal, while the mean of “ll-p” is slightly higher than the mean of “ll-np”. This can result from partial guessing. The means of “-np” groups in comparison with “-p” groups are equal (m(“hl-np”) = m(“hl-p”)) or less (m(“ll-np”) < m(“ll-p”)), but the std. deviation of the same groups are greater (s(“hl-np”) > s(“hl-p”)) or equal (s(“ll-np”) = s(“ll-p”)).

The revealed equivalence for both “-p” and “-np” groups in testing LOTS items indicates that there is no impact of penalty announcement on the balance between the explored latent variables: partial guessing and collaborative learning (a null hypothesis).

Varying the tests difficulty form LOTS to HOTS allows to underline the difference in the control groups behavior. Moving on now to support the alternative hypotheses.

3.3. Hypotheses H1 and H2

Let us look at the boxplots in Fig. 3a that present the LOTS test results. The boxplot of “hl-p” is comparatively short that suggests the students have a high level of agreement with each other. Then, we assumed that the boxplot of “hl-np” includes some portion of partial knowledge as it seems to be taller. However, the distribution is still symmetrical and, thus, the observations is evenly split at the median. The same values of the means show that many students have similar views on certain questions, but when there is no penalty for guessing, the students are more variable in their views. In contrast, both the boxplots “ll-p” and “ll-np” are comparatively tall: the students hold quite different opinions about what the correct answer is. Moreover, the distributions are skewed left as most of the observations are concentrated on the high end of the scale. Therefore, even if these groups of students take less partial guesses compared to “hl-p” and “hl-np”, they still demonstrate some partial knowledge. Subsequently, we conclude that the lower the order of thinking skills is, the more partial guesses in test results are. H1 is supported.

We can now return to Table 3. The HOTS test results indicate that the means of “-np” groups compared to “-p” groups are greater (m(“hl-np”) >m(“hl-p”)), (m(“hl-np”) >m(“hl-p”)) as well as the std. deviation of these groups (s(“hl-np”) >s(“hl-p”)), (s(“ll-np”) >s(“ll-p”)). This can be also explained either by the influence of guessing or the impact of collaboration. The boxplots in Fig. 3b may help us to understand these results better. Both the boxplots of “hl-p” and “hl-np” are skewed right because most of the observations are concentrated on the low end of the scale. Moreover, the boxplot of “hl-np” is much taller, so the students hold more different opinions about the correct answer. On the contrary, the boxplots of “ll-p” and “ll-np” are skewed left. Consequently, the observations are closer to the high end of the scale. The boxplot of “ll-np” is also taller, but to a lesser extent. Analyzing the results in Table 2, we suggested that the students when answering the HOTS tests may not have had enough knowledge, so they started working in collaboration to deepen their understanding. Although the students learned from collaboration (see the upper quartile of “hl-np” boxplot), some of them took pure guesses to get a correct answer (see the lower quartile of “hl-np” boxplot). Both “ll-” boxplots are skewed left as these students had more different incorrect opinions which can be interpreted as pure guesses. Hence, the higher the order of thinking skills is, the more pure guesses in test results are. H2 is supported.

3.4. Hypotheses H3 and H4

As follows from the discussed results, the balance between the degree of guessing and the impact of collaboration may depend on whether the penalty was announced, or the students had no penalty. We looked into this issue in more detail.

We combined the groups according to the students' level of knowledge and the penalty announcement. Table 4 shows the statistical results of the combined groups: “hl-”,“ll-”,“-p”,“-np”. We can see that all the results are statistically significant. For the LOTS tests, the mean of the group “-p” is slightly higher than the mean of “-np”. The same conclusion can be drawn with regard to the values of std. deviation. If we look at Fig. 4a, we see the boxplots for the LOTS test results. Although these boxplots are somewhat tall, they seem to be almost symmetrical. Moreover, we can see little skewness to the right as a result of the students' partial guessing (see H1). We concluded that testing questions that require lower-order thinking skills promotes learning from collaboration to a lesser extent in both cases: with penalty and no penalty for guessing. H3 is supported.

Analyzing the same results for HOTS (see Table 4, Fig. 4b), we came to the conclusion that both the boxplots are comparatively short with almost symmetrical distribution. This means that the students have a higher level of agreement with each other. Moreover, the students who had no penalty were able to reach common ground more effectively as the mean of the “-np” compared to “-p” is greater while the value of std. deviation is lesser (see Table 4). Consequently, testing questions that require higher-order thinking skills promotes learning from collaboration to a greater extent if there is no penalty for guessing. H4 is supported.

3.5. Hypotheses H5 and H6

Finally, we studied the test results from the viewpoint of the levels of students. Proving the hypotheses H1–H4, we paid careful attention to significant differences in the test results with respect to the level of students. The major difference is the following: even working in collaboration, “ll-” groups in contrast to “hl-” groups face difficulties in producing constructive ideas to get the correct

Table 4
Means, standard deviations, and *t*-test results subject to the order of thinking skills for combined groups (penalty announcement and knowledge level).

Group	Descriptive statistics				*t*-test			
	LOTS		HOTS		LOTS		HOTS	
	Mean	Std.dev.	Mean	Std.dev.	t	*p*-value	t	*p*-value
hl- ($K = 36$)	0.63	0.06	0.36	0.13	20.72	0.0002***	6.24	0.0034**
ll- ($K = 48$)	0.36	0.16	0.13	0.08	4.41	0.0216*	3.97	0.0166*
-p ($K = 44$)	0.50	0.10	0.22	0.07	10.16	0.002**	6.60	0.0027**
-np ($K = 40$)	0.49	0.12	0.27	0.09	8.33	0.0036**	7.07	0.0021**

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

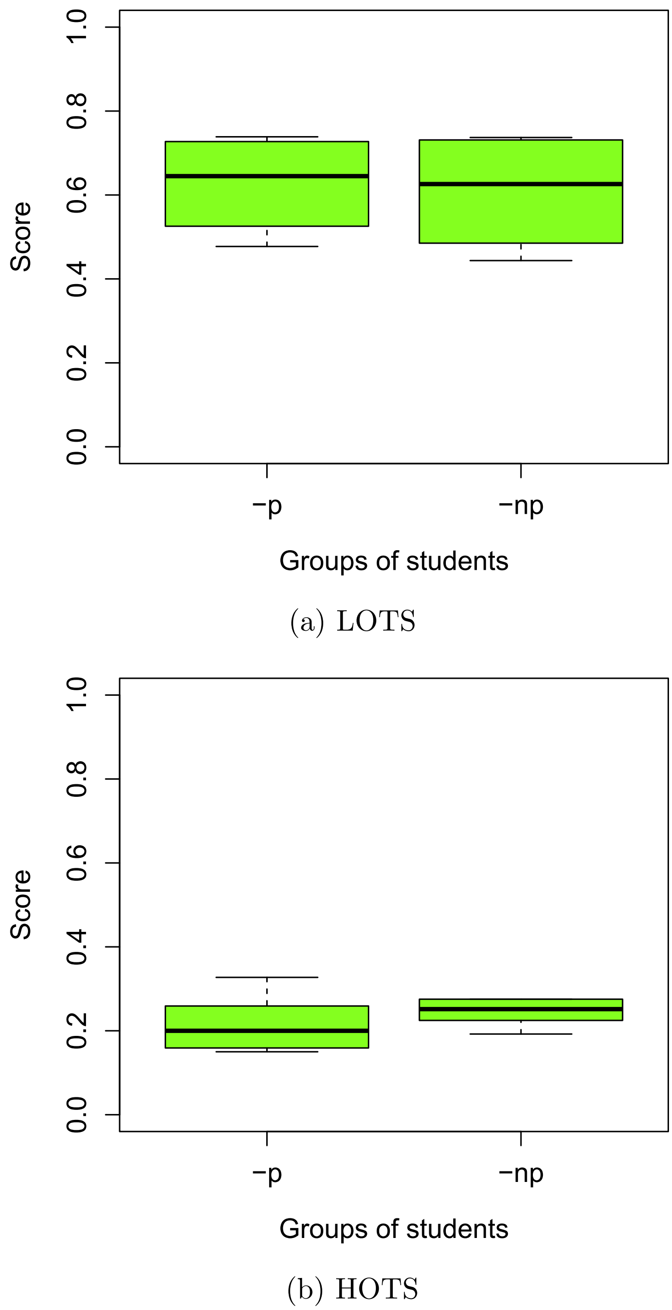

Fig. 4. Boxplots with regard to the order of thinking skills for combined groups (penalty announcement): “-p”, “-np”.

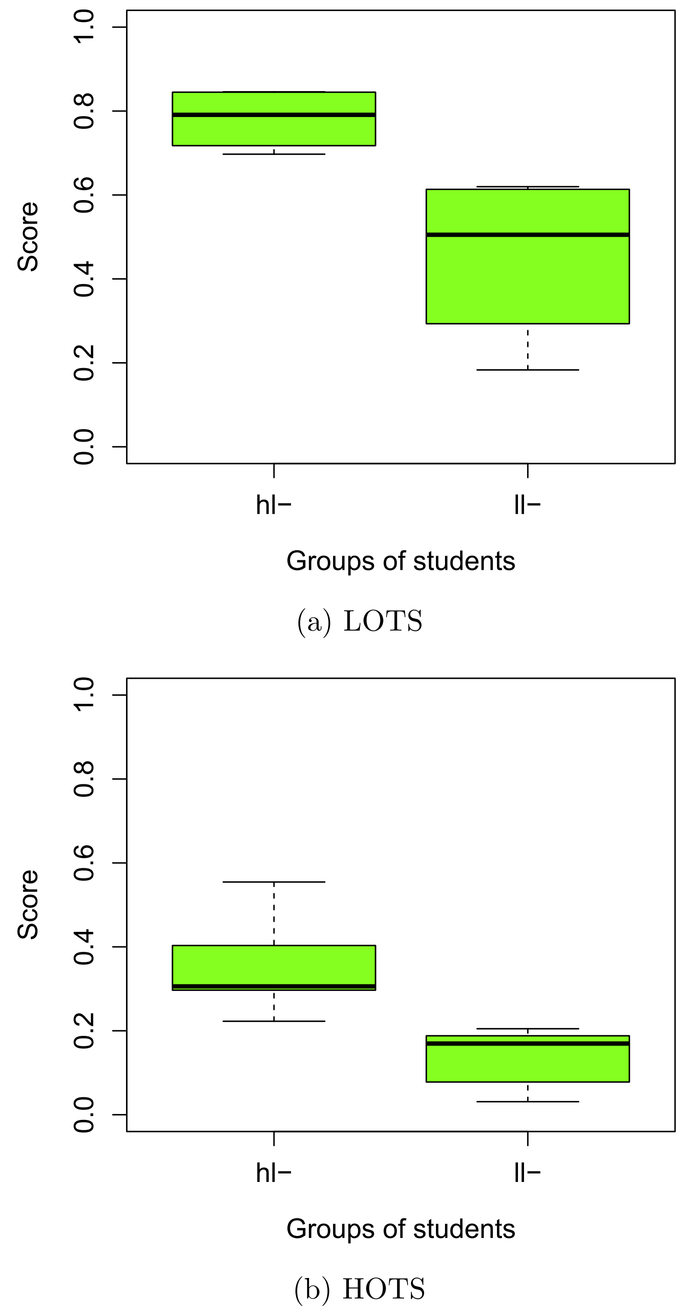

Fig. 5. Boxplots with regard to the order of thinking skills for combined groups (knowledge level): “hl-”, “ll-”.

answer. Moreover, as we noticed before, the lower p-value for “ll-” groups (see Tables 3 and 4) indicates more pure guessing and less stronger impact of collaboration.

We can observe this difference in Fig. 5a for the LOTS tests results and in Fig. 5b for the HOTS tests results. According to H3, there is a slight impact of collaboration presented in Fig. 5a. But we can see the result of taking more partial guesses by “hl-” groups - the boxplot is short and a little left skewed, and more pure guesses by “ll-” groups - the boxplot is much taller and significantly skewed to the left. Following the hypothesis H4, we now look at Fig. 5b that presents the tests results for HOTS. Here we can see that the boxplot of “hl-” groups is much higher than the same plot of “ll-” groups. Additionally, the first boxplot is dramatically right skewed (or positively skewed) while the other is drastically left skewed (or negatively skewed). This means that lower level students make more pure guesses and learn from collaboration to a lesser extent - that confirms H5. On the other hand, higher level students make more partial guesses and learn from collaboration to a greater extent. We concluded the validity of H6.

4. Discussion

The present results fully comply with the previous studies that stated positive findings for clicker activities through students' collaboration (McDonough & Foote, 2015). McDonough and Foote (2015) examined the impact of individual and shared clicker use on students' collaborative learning. The research results reported that shared clicker activities produce more collaborative reasoning than individual clicker activities. Moreover, shared clickers elicited better learning outcomes though it may depend on “the students' perception about having a “good” partner” (p. 244). As implications for further consideration, McDonough and Foote (2015) indicated the necessity of identifying clicker techniques as a matter of maximizing their benefits and, in this light, pointed up the importance of analyzing question difficulty that may stimulate students' collaboration and engagement. The present study is an attempt to fill this gap, examining how the level of thinking skills, related to question difficulty, and the level of shared partial knowledge affect the degree of guessing in answering the clicker questions.

Considering the penalty announcement as an external motivation tool may draw another meaningful comparison. Wang et al. (2017) have explored the impact of reward announcement on self-efficacy, self-regulated learning, motivation for achievement, and critical thinking. The findings of the research revealed the advantages unrewarded competition over rewarded competition in promoting students' collaboration. The present study has pointed to a similar conclusion: the announcement of penalty has a negative effect on promoting collaborative learning even if it leads to reducing pure guesses in test results.

5. Limitations and future directions

There are tree main limitations that need to be addressed regarding this research before broad generalizations may be made. First, the study presents a restricted sample of university students majored in engineering sciences. Second, even if the collected small sample helped us to support the hypothesis providing statistically significant estimates, a larger sample size would be more reliable. Finally, the groups categorization was non-blinded. But estimates' bias may be significantly reduced by keeping participants unaware of the goals and hypotheses of the study. In attempt to overcome the last limitation, in line with the blindness of procedure, we suggest the following directions for future research.

Firstly, the proposed models consider only two arbitrary points on the thinking skills line. The first point illustrates the combination of the questions that require LOTS, while the other - all the questions that require HOTS. Future research should take into consideration more points on the thinking skills line. That would allow to discuss the relationship between the type of thinking skills and the level of partial guessing. The results of this analysis could be used to better regulate guessing in a collaborative setting. Another promising direction for further research is the extension of the proposed models, which would include not only “footprints” of collaboration on the guessing line, but also “footprints” of guessing on the thinking skills line.

6. Conclusions

The present study was aimed at analyzing the problem of guessing regulation to promote collaborative learning. For this purpose, we tested two control groups of students “hl-” and “ll-”. They were asked to answer clicker questions, which increased in difficulty and were required the application of both LOTS and HOTS. To investigate the relationship between the influence of guessing and the impact of collaboration, two subgroups of students (“hl-p” and “ll-p”) were announced penalty while the other subgroups (“hl-np” and “ll-np”) had no penalty. In addition, all the subgroups were invited to collaborate.

We posed the research question on how the levels of cognition and the students' levels of knowledge influence on the degree of partial guessing in the collaborative learning environment. To answer this question, we developed two research models. Firstly, these models demonstrate the relationship between the order of thinking skills and the degree of guessing. Secondly, we consider that relationship with regard to the penalty announcement and on the hypothesis that students, working in groups, may be open to collaboration to a different extent. Finally, the developed models present the result complex relationship from the viewpoint of the levels of tested students.

To validate the research models, we put forward a set of hypotheses, which were proved by the results of statistical experiments. Based on these results, we came to the following conclusions: a) the announcement of penalty has a negative effect on promoting collaborative learning even if it leads to reducing pure guesses in test results; b) questions that require higher-order thinking skills promote collaborative learning to a greater extent; c) creating mixed level groups of students seems advisable to enhance learning from collaboration and, thus, to decrease the degree of pure guessing.

Acknowledgements

This work was supported by the Ministry of Education and Science of the Russian Federation, grant 074-U01. The authors would like to thank the reviewers for the valuable comments and suggestions. We are also grateful to Raoul Kessels for sharing the TEXcode (Kessels, 2012) used to draw the emoticons in Fig. 2.

Appendix. Tests instructions.

test	cognitive level	instruction
1	1 Know	Choose the correct preposition in brackets, if necessary. (5 items)
2	2 Understand	In the text field below, write the numbers of the sentences in which there is no preposition. (5 items)
3	3 Apply	Choose appropriate prepositions from the list below to complete the sentences. (5 items)
4	3 Apply	Complete the sentences with correct prepositions if necessary. (5 items)
5	4 Analyse	Correct mistakes in prepositions if necessary. (5 items)
6	4 Analyse	Fill in the blanks using the best choice(s) from the suggested lists. (5 items)
7	5 Evaluate	Reorder the sentences on your right so that they have the same preposition in each line with the sentences on the left. (5 items)
8	5 Evaluate	What picture best describes the sentence “The cat jumped over the wall.”? (4 items)
9	6 Create	Read the text below. Paraphrase each preposition OVER, using a phrase or a word, which is similar in meaning. (4 items)

References

Anderson, L. W. (Ed.), Krathwohl, D. R. (Ed.), Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (complete edition). New York: Longman.

Aluisio, S. M., Aquino, V. T., Pizzirani, R., & de Oliveira, O. N., Jr. (2003). Assessing high-order skills with partial knowledge evaluation: Lessons learned from using a computer-based proficiency test of English for academic purposes. *Journal of Information Technology Education, 2*, 185–201.

Arnold, J. C., & Arnold, P. L. (1970). On scoring multiple choice exams allowing for partial knowledge. *The Journal of Experimental Education, 39*(1), 8–13.

Barroso-Mendez, M. J., Galera-Casquet, C., Seitanidi, M. M., & Valero-Amaro, V. (2016). Cross-sector social partnership success: A process perspective on the role of relational factors. *European Management Journal, 34*(6), 674–685.

Basu, S. (2001). Improved small sample inference procedures for epidemiological parameters under cross-sectional sampling. *Journal of the Royal Statistical*

Society: Series D (The Statistician), 50(3), 309–319.
Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement, 21*(1), 65–88.
Bereby-Meyer, Y., Meyer, J., & Budescu, D. V. (2003). Decision making under internal uncertainty: The case of multiple-choice tests with different scoring rules. *Acta Psychologica, 112*, 207–220.
Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desired difficulty in the classroom. *Journal of Applied Research in Memory and Cognition, 3*(3), 165–170.
Bjork, E. L., Soderstrom, N. C., & Little, J. L. (2015). Can multiple-choice testing induce desirable difficulties? Evidence from the laboratory and the classroom. *The American Journal of Psychology, 128*(2), 229–239.
Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Javier Sese, F. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education, 62*, 102–110.
Bode, M., Drane, D., Ben-David Kolikant, Y., & Schuller, M. (2009). A clicker approach to teaching calculus. *Notices of the American Mathematical Society, 56*(2), 253–256.
Brady, M., Seli, H., & Rosenthal, J. (2013). "Clickers" and metacognition: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education, 65*, 56–63.
Bruff, D. (2009). *Teaching with classroom response systems: Creating active learning environments*. San Francisco, CA: Jossey-Bass.
Bustince, H., Fernandez, J., Hagras, H., Pagola, M., & Barrenechea, E. (2014). Interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: Towards a wider view on their relationship. *IEEE Transactions on Fuzzy Systems*. http://dx.doi.org/10.1109/TFUZZ.2014.2362149.
Butler, A., & Roediger, H. L., III (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.
Chan, C. K. K. (2012). Co-regulation of learning in computer-supported collaborative learning environment: A discussion. *Metacognition and Learning, 7*(1), 63–73.
Chien, Y.-T., Chang, Y.-H., & Chang, C.-Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review, 17*, 1–18.
Clark, H. H. (1968). On the use and meaning of prepositions. *Journal of Verbal Learning and Verbal Behavior, 7*, 421–431.
Cook, R., & Calkins, S. (2013). More than recall and opinion: Using clickers to promote complex thinking. *Journal on Excellence in College Teaching, 24*(2), 51–76.
De Backer, L., Van Keer, H., Moerkerke, B., & Valcke, M. (2016). Examining evolutions in the adoption of metacognitive regulation in reciprocal peer tutoring groups. *Metacognition and Learning.*. http://dx.doi.org/10.1007/s11409-015-9141-7.
De Backer, L., Van Keer, H., & Valcke, M. (2014). Socially shared metacognitive regulation during reciprocal peer tutoring: Identifying its relationship with students' content processing and transactive discussions. *Intructional Science, 43*(3), 323–344.
Dehnad, A., Nasser, H., & Hosseini, A. F. (2014). A comparison between three and four option multiple choice questions. *Procedia - Social and Behavioral Sciences, 98*, 398–403.
Dupuis, M., Khadeer, S., & Huang, J. (2017). "I Got the Job!": An exploratory study examining the psychological factors related to status updates on Facebook. *Computers in Human Behavior, 73*, 132–140.
Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: David McKay.
Espinosa, M. P., & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology, 54*, 415–425.
Ghafournia, N. (2013). The relationship between using multiple-choice test-taking strategies and general language proficiency level. *Procedia - Social and Behavioral Sciences, 70*, 90–94.
Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and Instruction, 22*(6), 401–412.
Gucht, F. V., Willems, K., & Cuypere, L. (2007). The iconicity of embodied meaning. Polysemy of spatial prepositions in the cognitive framework. *Language Sciences, 29*, 733–754.
Huff, K. C. (2015). The comparison of mobile devices to computers for web-based assessments. *Computers in Human Behavior, 49*, 208–212.
International language standards. Cambridge English. (2016). *Cambridgeenglish.org*. http://www.cambridgeenglish.org/cefr/ (Accessed 21 May 2016).
Isohatala, J., Jarvenoja, H., & Jarvela, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research, 81*, 11–24.
Jarvela, S., & Hadwin, A. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39.
Jarvela, S., & Hadwin, A. (2015). Promoting and researching adaptive regulation: New Frontiers for CSCL research. *Computers in Human Behavior, 52*, 559–561.
Jarvela, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction, 42*, 1–11.
Jimánez Catalán, R. M. (1996). Frequency and variability in errors in the use of English prepositions. *A Journal of English and American Studies, 17*, 171–188.
Kemmerer, D. (2005). The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia, 43*, 797–806.
Kessels, R. (2012). *TEXample.net}. TikZ examples. Example: Emoticons*. http://www.texample.net/tikz/examples/emoticons/ (Accessed 01 May 2016).
Kleiner, L. F. (2005). The semantics of English prepositions. Book review. *Journal of Pragmatics, 37*, 775–779.
Kubinger, K. D., Holocher-Ertl, S., Reif, M., Hohensinn, C., & Frebort, M. (2010). On minimizing guessing effects on multiple-choice items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. *International Journal of Selection and Assessment, 18*, 111–115.
Lantz, M. E. (2010). The use of 'Clickers' in the classroom: Teaching innovation or merely an amusing novelty. *Computers in Human Behavior, 26*(4), 556–561.
Lantz, M. E., & Stawiski, A. (2014). Effectiveness of clickers: Effect of feedback and the timing of questions on learning. *Computers in Human Behavior, 31*, 280–286.
van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of multiple computer-supported collaborative groups. *Computers in Human Behavior, 52*, 233–242.
Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education - is it still a matter of number right scoring or negative making? *Studies in Educational Evaluation, 39*, 188–193.
Levesque, A. A. (2011). Using clickers to facilitate development of problem-solving skills. *CBE Life Sciences Education, 10*(4), 406–417.
Lindstromberg, S. (2010). *English prepositions explained* (revised ed.). Amsterdam: John Benjamins.
Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43*, 14–26.
Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44*, 1085–1101.
Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Lawrence: Erlbaum Associates, Inc.
Lo, J.-J., Wang, H.-M., & Yeh, S.-W. (2004). Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia. *Computers & Education, 42*, 45–63.
Malmberg, J., Jarvela, S., Jarvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior, 52*, 562–572.
Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem-solving. *Computers in Human Behavior, 18*, 623–632.
Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*, 51–57.
McDonough, K., & Foote, J. A. (2015). The impact of individual and shared clicker use on students' collaborative learning. *Computers & Education, 86*, 236–249.
Mueller, C. M. (2011). English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? *System, 39*, 480–490.
Navimipour, N. J., & Zareie, B. (2015). A model for assessing the impact of e-learning systems on employees' satisfaction. *Computers in Human Behavior, 53*, 475–485.
Ngu, B. H., & Rethinasamy, S. (2006). Evaluating a CALL software on the learning of English prepositions. *Computers & Education, 47*, 41–55.
Nickerson, R. S., Butler, S. F., & Carlin, M. T. (2015). Knowledge assessment: Squeezing information from multiple-choice testing. *Journal of Experimental Psychology: Applied, 21*(2), 167–177.
Nicol, D. (2007). E-assessment by design: Using multiple-choice tests to good effect. *Journal of Further and Higher Education, 31*(1), 53–64.
Raes, A., Schellens, T., De Wever, B., & Benoit, D. F. (2016). Promoting metacognitive regulation through collaborative problem solving on the web: When scripting does not work. *Computers in Human Behavior, 58*, 325–342.
Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal, 34*(6), 598–605.
Ronkko, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat Emptor. *Personality and Individual Differences, 87*, 76–84.
Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., et al. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*, 122–124.
Stowell, J. R. (2015). Use of clickers vs. mobile devices for classroom polling. *Computers & Education, 82*, 329–334.
Streukens, S., & Leroi-Werelds, S. (2016). Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. *European Management Journal, 34*(6), 618–632.
Tarrant, M., & Ware, J. (2010). A comparison of the psychometric properties of three- and four-option multiple-choice questions in nursing assessments. *Nurse Education Today, 30*, 539–543.
Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers & Education, 34*, 37–49.
Tyler, A., & Evans, V. (2003). *The semantics of English prepositions*. Cambridge: Cambridge University Press.
Valaei, N., & Baroto, M. B. (2017). Modelling continuance intention of citizens in government Facebook page: A complementary PLS approach. *Computers in Human Behavior*. http://dx.doi.org/10.1016/j.chb.2017.03.047.
Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education, 54*, 455–461.
Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2011). Comparison of oral

examination and electronic examination using paired multiple-choice questions. *Computers & Education, 56*, 616–624.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Mass.: Harvard University Press.
Wang, X., Wallace, M. P., & Wang, Q. (2017). Rewarded and unrewarded competition in a CSCL environment: A coopetition design with a social cognitive perspective using PLS-SEM analyses. *Computers in Human Behavior, 72*, 140–151.
Willaby, H. W., Costa, D. S. J., Burns, B. D., MacCann, C., & Roberts, R. D. (2015). Testing complex models with small sample sizes: A historical overview and empirical demonstration of what partial least squares (PLS) can offer differential psychology. *Personality and Individual Differences, 84*, 73–78.
Winne, P. H. (2015). What is the state of the art in self-, co- and socially shared regulation in CSCL? *Computers in Human Behavior, 52*, 628–631.
Zareva, A. (2012). Partial word knowledge: Frontier words in the L2 mental lexicon. *International Review of Applied Linguistics in Language Teaching, 50*(4), 277–301.