CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Assessing efficiency of prompts based on earner characteristics
Authors	Backhaus, Joy;Jeske, Debora;Poinstingl, Herbert;Koenig, Sarah
Publication date	2017-02-10
Original Citation	Backhaus, J., Jeske, D., Poinstingl, H. and Koenig, S. (2017) 'Assessing Efficiency of Prompts Based on Learner Characteristics', Computers, 6(1), 7. doi: 10.3390/computers6010007
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.3390/computers6010007
Rights	© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). - http://creativecommons.org/licenses/by/4.0/
Download date	2026-11-15 02:26:41
Item downloaded from	https://hdl.handle.net/10468/4082



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

computers

MDPI

Article

Assessing Efficiency of Prompts Based on Learner Characteristics

Joy Backhaus [1,*], Debora Jeske [2], Herbert Poinstingl [3] and Sarah Koenig [1]

1 Medical Teaching and Medical Education Research, University Hospital Wuerzburg, 97080 Wuerzburg, Germany; koenig_sarah@ukw.de
2 School of Applied Psychology, University College Cork, T23 K208 Cork, Ireland; d.jeske@ucc.ie
3 Faculty of Psychology, University of Vienna, A-1010 Vienna, Austria; herbert.poinstingl@gmail.com
* Correspondence: Backhaus_J@ukw.de; Tel.: +49-931-201-55209

Academic Editors: Marko Tkalčič, Berardina Nadja De Carolis, Marco de Gemmis and Andrej Košir
Received: 28 November 2016; Accepted: 7 February 2017; Published: 10 February 2017

Abstract: Personalized prompting research has shown the significant learning benefit of prompting. The current paper outlines and examines a personalized prompting approach aimed at eliminating performance differences on the basis of a number of learner characteristics (capturing learning strategies and traits). The learner characteristics of interest were the need for cognition, work effort, computer self-efficacy, the use of surface learning, and the learner's confidence in their learning. The approach was tested in two e-modules, using similar assessment forms (experimental $n = 413$; control group $n = 243$). Several prompts which corresponded to the learner characteristics were implemented, including an explanation prompt, a motivation prompt, a strategy prompt, and an assessment prompt. All learning characteristics were significant correlates of at least one of the outcome measures (test performance, errors, and omissions). However, only the assessment prompt increased test performance. On this basis, and drawing upon the testing effect, this prompt may be a particularly promising option to increase performance in e-learning and similar personalized systems.

Keywords: e-learning; assessment; prompting; personalization; self-regulation

1. Introduction

One of the challenges in personalized e-learning initiatives pertains to the selection and consideration of tools based on learner characteristics that may increase performance for different types of learners [1]. To date, many user and learner models personalize the learning experience by implementing tutoring support based on perceived or self-reported learner characteristics [2,3]. Prompts can help learners to engage more with the material but also reflect on their learning progress [4]. They may also remind learners of appropriate learning and questioning strategies, which in turn improve performance in an online environment [5].

A variety of learner characteristics are relevant in self-regulation [6], specifically those that influence "cognitive, metacognitive, motivational, volitional and behavioral processes." ([6], p. 194). Several learner characteristics were considered important for self-regulation in the context of this study. These included five learner characteristics believed to be relevant for the implementation of prompts: willingness to commit to making an effort, need for cognition, processing strategies (surface strategy), computer self-efficacy and confidence. All of which can be considered traits rather than states since they are stable characteristics, which are unlikely to change frequently. In the next section the five learner characteristics on which prompts are based are presented. The assessment of these learner characteristics is referred to as the "learner profile" in the following sections.

Work effort: The first learner characteristic of interest is work effort, namely the extent to which individuals, in order to bring tasks to an end, persist, and invest further effort when they face difficulties [7]. Greater work effort sets the stage for better performance.

Need for Cognition: Another relevant personality characteristic is the learner's need for cognition. This is defined as the inherent need to make sense of situations by cognitively structuring them. Greater need for cognition means the learner is motivated to try to understand new information [8], while those with lower scores are not driven to learn for the sake of gaining new insight—in which case a prompt may be helpful.

Surface Strategy: Behavioral processes come into play in terms of the strategies which capture the different ways in which learners approach a task, such as surface vs. deep processing strategies [9]. Those who engage in more superficial than in-depth strategies tend to roam through a text or e-module. This enables them to learn about the most obvious aspects and gain an overview. However, they may not actively improve their understanding. As a result, using this strategy is often negatively related to performance [10].

Computer self-efficacy: Finally, behavioral processes may also reflect learners' prior experience as well as perceived capability to learn materials effectively when they use electronic devices and e-learning platforms. This is in line with past research which has shown that computer self-efficacy is a significant predictor of performance [11].

Confidence: The overall confidence of learners may also play a role in terms of how they progress when they learn. The relevance of assessing confidence in the context of e-learning is important for three reasons. First, when predicting performance on a test, individuals include estimates of retention. Poorer performers do so less well than good performers [12]; Second, many individuals do not adjust their performance judgments upward or downward enough to account for differences in their ability. In addition, they do not accurately estimate changes in mean performance related to task difficulty [13]; Third, the presentation of a metacognitive prompt may reduce the illusion of understanding that may arise in learners [14].

The goal of the current study was to outline and examine a personalized prompting approach aimed at eliminating performance differences. The next section discusses the use of prompts as tutoring aids for self-regulation and the link to work effort, need for cognition, surface strategy, computer self-efficacy and confidence. This section also provides the basis for our introduction to our own prompting approach.

1.1. Prompting for Improved Self-Regulation

In order to facilitate learning, several different and very specific prompts may be employed that focus on problem-solving [15] general instruction [5], self-explanation [16], or encouraging recall, reasoning, or observation [5]. Prompts can serve multiple functions and can focus on cognitive learning processes, metacognitive self-regulation, and resource management [17]. Many of these prompts are basically scaffolds. They enable learners to make sense of what is required of them when solving a problem. In the domain of science education, a number of reviews exist that provide further insight (e.g., [18]). One study is particularly relevant here. Moreno, Mayer, Spires, and Lester employed a prompt that required students to feedback how much the learning material presented to them also helped them understand the link between plant design and the environment [19]. Students were subsequently prompted to recall what is required of them and monitor what they have learned [4], demonstrating how prompts may support self-regulation.

The central aim in the current research was to increase performance through the careful design and implementation of prompts, either by directly targeting maladaptive self-regulatory behavior or by increasing the accuracy of learner's self-assessments of their learning progress to date (via a confidence assessment). These aims are approached by implementing five prompts which are presented next. Instructors and e-learning designers alike aim to increase performance among those learners who generally tend to perform less well—at least as noted in previous performance

records. Two potential options are available. For example, a *strategy prompt* can explain strategies to participants, including examples on how they may succeed. This may also improve self-regulation and, thus, improve performance [20].

While strategy prompts are useful means to influence the learning process, *assessment prompts* can prompt learners into action by having them provide evidence of learning, increase self-evaluation and by this mean make use of the testing effect [21]. The testing effect simply describes the phenomenon that learners are able to recall information better if they have been tested on it [22]. Challenging students to recall what they have learned is important since many learners are not effective self-monitors and, therefore, often overestimate their learning success [23]. These activities change the learner's self-evaluation of progress made, leading to more accurate judgments [24] (experiment 4). Self-evaluations have been shown to increase performance in computer-supported learning [25].

Reducing overconfidence in those learners that are more likely to be inaccurate self-assessors is another challenge. Overconfidence is particularly an issue for over-confident individuals who actually have lower ability and performance [13,26], as such overconfidence may also affect progression and effort. A *verification prompt* can be useful here as it may reduce over-confidence or faking (reflected in extremely high scores on certain learning characteristics, such as work effort and need for cognition), as well as assist self-regulation in online settings [26,27]. In addition, confidence assessments during the learning process (in form of a *confidence prompt*) may also reduce overly optimistic learner expectations regarding their learning progress [28]. On the other hand, alternative prompts can be used to address underconfidence in learners (even when they actually the ability to do well). Underconfidence, like overconfidence, can have a significant effect, as this may lead to participants to disengage earlier or drop out. An *explanation prompt* can be helpful to ensure that participants understand the purpose and relevance of the e-module. Such a prompt may potentially reduce concerns among those learners who are less confident and indirectly help them to persist and maintain their willingness to invest efforts during the learning process [20]. In addition, a *motivation prompt* can provide learners with information designed for different levels of user competence to encourage completion, particularly when this process is both lengthy and effortful for learners [29].

1.2. Current Study Rationale

Outcomes of e-learning design are often based on comparing specific groups of recipients. We believe that more work is needed to evaluate prompt design and effectiveness by considering not just comparative samples, but also matched participant design. Based on the learner profile a matched case-control design was utilized. Participants were of the experimental group were matched to those participants of the control group, who had identical values on the learner profile. This methodological design enables to evaluate subjects with equivalent scores on the learner profiles perform different when prompted.

The current study will address two research questions (RQ):

- RQ1: Do participants perform better when prompted compared to a participants with an identical learner profile when not prompted?
- RQ2: In case the prompts do not lead to better performance, can more efficient threshold be determined in terms of performance outcome?

By these means we intend to determine an individual prompting approach tailored to address characteristics relevant to learning and subsequently performance. Since the novelty of the work lies in the presentation of prompts based on cut-off values derived from self-reported data, explicit recommendations for threshold refinement are deduced. The latter is exploratory in nature since no research work addressing efficiency of thresholds for a prompting approach based on a learner profile could be determined.

2. Design and Methods

The next two sections outline the design (including learning context, learning profile, and basis for prompts), followed by the methods (including a description of the recruitment procedure and participants and follow-up measures).

2.1. Design: Learning Context

The following section describes the learning context and the implementation (as well as content) of the prompts within the learning context. The next section describes the e-learning context in more detail.

The study involved two different e-modules of similar length, similar readability indices, the same learning profile, the same number of test questions, and follow-up queries. Both e-modules covered topics related to health at work and teamwork. The five chapters on team development describe the four stages of forming, storming, norming, and performing [30], which are likely to occur when a new team is assigned a task. The five chapters on shiftwork taught about health effects of shiftwork. Each chapter ended with a multiple choice test question. The mean time spent on the e-module where 9.9 min (*SD* = 4.7 min) and 9.5 min (*SD* = 5.23) for the e-module on shiftwork and team development respectively.

Descriptive statistics for the modules are presented in Table 1 including the Automated Readability Index (ARI) [31] and the Flesh Reading Ease Readability Formula (FRERF) [32]. The ARI and the FRERF are formulae to measure the understandability of texts. The ARI corresponds to the US grade level required to comprehend a text, an ARI score of "8" indicates that a normal seventh grader should have no difficulty understanding a text, whereas a value of "14" indicates that a text is most appropriate for college students. Higher values of the FRERF simply indicate that a text is easier to comprehend. Values from 0 to 30 indicate that a text is difficult and most appropriate for academics, whereas values close to "100" indicate that a text is very easy. The main difference between the indices is that the ARI is based on characters, words and sentences whereas the FRERF incorporates syllables. Since our main aim was to ensure comparability of text difficulty we decided for these rather simple formulae [33]. The ARI varies between 8.73 and 11.70, which indicates that college students should not have any difficulty understanding the texts. The latter is confirmed by the FRERF.

Table 1. Descriptives for the module on team development/shiftwork.

	WC	CH	SEN	CH/WC	WC/SEN	ARI	FRE
Chapter 1	206/231	1253/1132	16/17	6.08/4.90	12.88/13.59	13.66/8.45	35.90/63.00
Chapter 2	206/272	1087/1332	14/19	5.28/4.90	14.71/14.32	10.78/8.79	60.20/55.20
Chapter 3	182/268	944/1451	11/12	5.19/5.41	16.55/22.33	11.27/13.24	64.00/49.50
Chapter 4	177/223	963/1051	12/14	5.44/4.71	14.75/15.93	11.57/8.73	54.60/64.60
Chapter 5	170/293	921/1676	14/9	5.42/5.72	12.14/32.56	10.16/11.79	59.90/39.30

Note. WC = word count; CH = characters; SEN = sentences; CH/WC = ratio of characters to words; WC/SEN = ratio of word to sentences; ARI = Automated Readability Index; FRE = Flesh Reading Ease Readability Score.

2.2. Learning Profile

The e-modules each featured one learning profile at the beginning. The following measures were utilized in the learning profile.

2.2.1. Work Effort

Work effort was measured using five items [34], the first three measuring persistence, the last two measuring work effort intensity. An example item is "When I start an assignment I pursue it to the end". The response options range from (1) "strongly disagree" to (5) "strongly agree".

The scale utilized ranges from 5 to 25, whereas 20 is considered an extremely high work effort score. When the scores fall between 0 and 12, work effort is categorized as low.

2.2.2. Computer Self-Efficacy

Computer self-efficacy was measured using three items [35]. An example item is "I feel confident troubleshooting computer problems". The response options range from (1) "strongly disagree" to (5) "strongly agree". The minimum score is three, the maximum score is fifteen. When the scores fall between 0 and 12, computer self-efficacy is categorized as low.

2.2.3. Need for Cognition

Need for cognition was measured using five items [36]. An example item is "I would prefer complex to simple problems". The response scale ranged from (1) "extremely uncharacteristic" to (5) "extremely characteristic". The score ranges from 5 to 25. When the score ranges between 0 and 15 for women (and 0 to 18 for men), need for cognition is considered low.

2.2.4. Surface Strategy

Surface strategy was measured using three items of the two-factor Study Process Questionnaire [9]. An example item is: "In order to understand something, I tend to study more than what may be necessary". To avoid socially desirable responses all items were reverse-coded [37]. The Response options for all six items are a five-point Likert scale ranging from (1) "never or only rarely true of me" to (5) "always or almost always true of me". The score ranges from 5 to 15 points maximum.

High surface strategy is associated with scores between 13 and 15.

2.2.5. Confidence Assessment

All participants were asked how confident they are that they could answer a test questions about the previous sections. The response options can be selected on a visual analogue scale ranging from 0% to 100%.

2.3. Prompts: Design, Thresholds and Implementation

The prompts and the specific instructions for each are presented next. The *verification prompt* instructed learners as follows: "You have reported an exceptionally high score on e.g., work effort. Please confirm this is the score you meant to indicate!" The *explanation prompt* stated "This e-module is an important tool for you to familiarize yourself with the topic of team development/effects of shift work on health. Completing this e-module attentively and your own pace can provide you with the skills and knowledge to succeed at work and in subsequent learning units." The *strategy prompt* outlined: "The following strategies may help you succeed. Remember to check that you have understood the concepts covered so far. To help you, try to summarize, write down important concepts and re-visit difficult parts. When you have read and understood these instructions, please click "ok" to proceed." The *motivation prompt* reminded learners that "The e-module was designed for all kinds of learners with different skills and backgrounds. You can improve your performance by remembering your goals for taking this e-module. Focus on these goals for taking this e-module and how you have successfully learned in the past. Please feel assured that with appropriate effort, all participants should be able to complete the e-module successfully. Your hard work will pay off!" The *assessment prompt* told learners that "In order to assess progress, we would hereby like you to enter five key words that summarized what you have learned so far in this e-module. This small assessment serves to ensure that you as a learner can assess your own progress successfully." Figure 1 shows how the prompts were positioned in each e-module in relation to the chapters. Table 2 presents an overview of prompts, the learning characteristics of interest and thresholds used to trigger prompts. Thresholds were derived from data obtained in previous studies by the authors [38–40]. We found that women tend to have

lower judgment of learning scores overall. In order to determine when to trigger encouraging prompts, we decided to lower the confidence threshold for women by about 3%–5% compared to men, who rate their confidence on average as higher than women. Furthermore we found significant age differences regarding judgment of learning in several of our preliminary studies. Older participants reported being more confident about their learning than any younger group of participants. After correcting their confidence downward, they nevertheless continued to be over-confident as indicated by the (difference) scores produced between confidence and test items. Gender differences in self-rating have also been reported for need for cognition [41]. Subsequently, thresholds were adapted to age and gender for these scales.

Table 2. Prompt set up.

Variable	Thresholds for Prompts	Prompts
Need for cognition	Low (score f $\leq$ 15; score m $\leq$ 18) High (score f $\geq$ 18; score m $\geq$ 20–24) Extremely high (25)	Strategy – Verification/Assessment prompt
Work effort	Low (score $\leq$ 12) High (score $\geq$ 16) Extremely high (=20)	Explanation – Verification
Computer self-efficacy	Low (score $\leq$ 12) High (score $\geq$ 13)	Motivation –
Surface strategy	Low (score $\leq$ 12) High (score $\geq$ 13)	– Strategy
Confidence assessment	Low (age up to 34 $\leq$ 75; age 35+ $\leq$ 78) High (score > 76–90)	Motivation –
(JOL)	Very high (up to 34 $\leq$ 91; 35+ $\leq$ 94)	Assessment prompt

Note. f = female; m = male. The above learning characteristics were based on their relevance to learning. The thresholds were based on previous research conducted by the authors using the same learning characteristics.

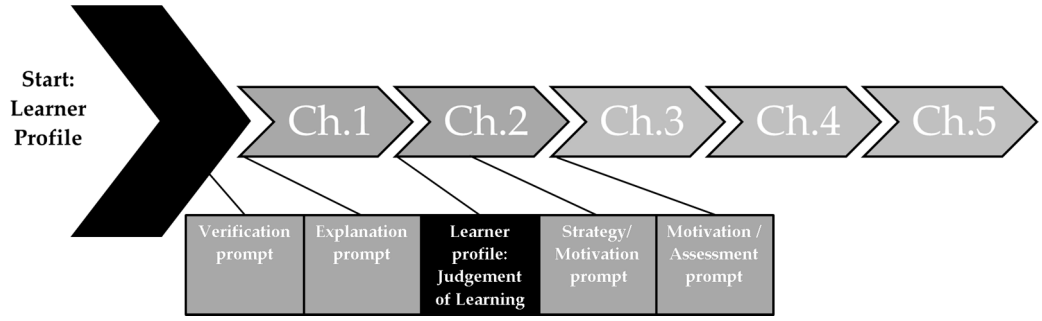


Figure 1. Positioning of prompts in both e-modules.

All participants in the experimental group faced one to three prompts maximum. Prompts could be triggered in a specific order (cf. Figure 1). The verification prompt is triggered upon completion of the learning profile (when individuals report extremely high work effort and need for cognition that may be indicative of faking or disengagement). The explanation prompt explains the purpose/relevance of the e-module and is situated at the beginning of the first chapter to participants with low work effort scores. The confidence prompt is presented at the end of the first chapter, and constitutes the basis for two further prompts: a strategy prompt and a motivation prompt (process prompts). The strategy prompt, if triggered, is presented in response to high surface strategy and low need for cognition. Alternatively, the motivation prompt is triggered by low confidence in learning or low computer self-efficacy. The assessment prompt is triggered in reaction to extremely

high confidence and high need for cognition reported in the learning profile. This prompt includes a request to participants to generate keywords and presented in response to possible overconfidence.

2.4. Methods: Procedure and Participants

All participants were recruited via their instructors at two Midwestern universities in the USA, a private English-speaking university in Germany and a university in the UK. Participation was voluntary. The mode of recruitment of participants was identical across both e-modules (announcement by instructor in the class). All participants were social science undergraduates who participated for extra credit. Once participants entered the online e-modules and read the study information, they could only proceed if they consent to participate. The learning profile was presented next (items assessed work effort, computer self-efficacy, need for cognition, surface strategy). Upon completion of the profile, participants were randomly allocated into the control group (no prompts) or the experimental groups (prompts triggered by learning profile) on an allocation rate of 1:1:1. All groups subsequently went through the e-module and completed five test sections, followed by a demographics section and a debrief statement.

As the primary goal was to determine if there were any module-independent effects of the prompts, the two datasets obtained from the two e-modules (n_1 = 225; n_2 = 273) were combined into one (N = 413). 243 participants (37%) belonged to the control group (not-prompted) and 413 (63%) to the experimental group (received a prompt based on their learning profile). Fifty participants were male (20.6%) and 193 female (79.4%) in the control group and aged 17 to 46 years old (M = 22.02, SD = 4.13). The experimental group included 107 male participants (25.9%) and 306 female (74.1%) with an average age range of 17 to 52 (M = 20.63, SD = 3.16). To facilitate readability the expression control and experimental group is used in throughout the following sections.

2.5. Methods: Outcome and Demographic Measures

A number of measures were collected during or at the end of the e-learning modules. These did not feature in the learning profile.

2.5.1. Test Performance

All participants were asked five multiple-choice questions embedded into the e-module, the maximum score participants could obtain were 15 points. Each chapter of the module featured one test question with one or more correct responses. Two test scores were created for each dataset. One value considered overall performance of participants. The corrected test performance was computed by deducting the number of errors made by the participants from their overall performance score. This step was important as participants were able to select all response options; not correcting for test error could skew the data and make individuals who were guessing appear more successful. Test performance was measured by correct responses to multiple-choice questions, errors, and omissions. In addition to errors the number of omissions was also recorded. Omissions are defined as correct response options participants did not select. Subsequently omissions are the reverse function of the test score.

2.5.2. Demographics

Demographical information included gender and age.

3. Results

The results section was split into three subsections. The subsections outline the general characteristics of the measures used, the evaluation of prompt effectiveness, and the specific prompt results.

3.1. Descriptives and Scale Performance for Both Conditions

Descriptives and scale descriptives for the control and experimental group are included separately (see Table 3).

Table 3. Descriptives and scale characteristics.

		Cronbach's α (Reliability Coefficient)		*M (SD)*	
Scale	**Items**	**Control**	**Experimental**	**Control**	**Experimental**
Need for cognition	5	0.74	0.76	3.50 (0.74)	3.41 (0.74)
Work effort	5	0.74	0.73	4.16 (0.55)	4.07 (0.56)
Computer self-efficacy	3	0.89	0.91	2.64 (0.87)	2.81 (0.96)
Surface learning	3	0.78	0.83	2.56 (1.03)	2.67 (1.09)
Judgement of Learning	1	n.a.	n.a.	71.23 (19.69)	69.48 (24.03)

Note. Items for surface learning were all reverse-coded when filling in learning profile. No remarkable skew or kurtosis (all values below +/1) except for work effort (kurtosis 1.46) occurred for experimental group. No remarkable skew or kurtosis (all values below +/1) values were noted for the control group. Since JOL is one item only, no α-value could be computed.

Scales performed similarly well in both subsamples. The two groups did not differ significantly in terms of their overall learning characteristics. As noted in Table 3, there were no significant differences for any of the learning characteristics between the control- and experimental group. In all, 173 (41.9%) of the participants in the experimental group received the strategy prompt, 62 (15%) participants received the assessment prompt, 23 (5.6%) participants received the explanation prompt, 37 (9%) participants received the verification prompt and 399 (96.6%) received the motivation prompt. The majority of learner characteristics correlated significantly with the test score in both conditions (Table 4).

Table 4. Correlation of scales and outcome measures for the experimental and the control group.

	Correlation Coefficient		
Scale	**Test Score**	**Errors**	**Corrected Test Score**
Need for cognition	0.057 (0.128 **)	0.047 (0.090)	0.020 (0.054)
Work effort	0.139 ** (0.180 **)	0.072 ^{t} (0.096)	0.075 (0.097 *)
Computer self-efficacy	−0.130 ** (−0.152 **)	0.000 (−0.028)	−0.113 ** (−0.116 *)
Surface learning	−0.058 (−0.137 **)	−0.087 * (−0.149 **)	0.005 (−0.024)
Confidence	0.255 ** (0.307 **)	−0.028 (−0.010)	0.240 **(0.277 **)

Note. ^{t} < 0.10; * $p < 0.05$; ** $p < 0.001$; $N = 656$. We used Spearman's rho to compute the correlation coefficients since data were not normally distributed. Results for the control group are in parenthesis, e.g., 0.057 is the correlation coefficient between need for cognition and test score for the experimental group, and 0.128 is the correlation coefficient between need for cognition and test score for the control group.

3.2. Analyses to Assess Prompt Effectiveness

The main goal of the current analysis was to assess whether receiving prompts at specific levels of learner characteristics will be effective as predicted. To respond to this question the following two steps of analysis were conducted. The first step consisted of a confirmatory analysis to assess in-between group comparisons for performance at the threshold-level proposed (cf. Table 1). Group comparisons between the experimental and control group were conducted using an ANOVA bootstrap approach [42]. This analysis examined whether specific learner characteristics (such as low need for cognition) result in underachievement if no prompt is provided among matched learners (e.g., learners in the control group who had the same scores in the learner profile). The second step of analysis evaluated the effectiveness of existing prompts and considered the possibility of identifying more efficient thresholds. This step enabled us to examine whether the threshold of prompts used in our study may need to be adjusted—presenting a starting point for recommendations for future studies.

The second step of the analysis was only conducted when no significant between-group differences were found in the first step of the analysis.

3.3. Results for Different Prompts

The following section presents the results for every prompt. Findings are summarized in Table 5.

3.3.1. Strategy Prompt

In the first step of the analysis, participants of the control group were matched to those of the experimental group concerning their values for need for cognition. No significant performance differences for female participants with a value $\leq$15, and males with a value $\leq$18, were determined (thresholds determined for prompts; $p > 0.05$). In other words, the performance of individuals with identical scores in the control and experimental group was not significantly different. This contradicts our expectations and suggests that the threshold chosen was not efficient for the population tested. The strategy prompt based on surface strategy, was only triggered two times. Subsequently, no valid group comparisons could be made. A lower threshold for the prompt has to be discussed for future implementations. In the second step of the analysis, we evaluated whether better thresholds for this prompt could be determined. Women who scored 15 and men who scored 17 on the need for cognition scale did indeed perform better than participants with identical scores in the control group.

Table 5. Prompt specific results and suggestions for future studies.

Prompt	Thresholds for Prompts	Better Performance	Suggested Threshold for Future Studies
Strategy	Low need for cognition (score w $\leq$ 15; score m $\leq$ 18)	No	We would suggest re-evaluating use.
	High surface strategy (score $\geq$ 13)	No	
Assessment	Very high judgment of learning (up to 34 $\leq$ 91; 35+ $\leq$ 94)	Yes	Prompt should be used in future studies using the thresholds proposed.
	Extremely high (25) need for cognition	Yes	
Explanation	Low work effort (score $\leq$ 12)	No	We would suggest re-evaluating use.
Verification	Extremely high (=20) work effort	No	We would suggest re-evaluating use.
	Extremely high (25) NFC	No	
Motivation	Low computer self-efficacy (score $\leq$ 12)	No	Recommendation: change threshold of <12 and smaller
	Low confidence (age up to 34 $\leq$ 75; age 35+ $\leq$ 78)	No	Recommendation: change threshold of <75 and smaller regardless of age

3.3.2. Assessment Prompt

The assessment prompt was triggered based on participants reporting very high judgment of learning levels ($\geq$91 and 94, respectively, for participants aged below and older than 34, see Table 1). At these thresholds significant differences emerged between the experimental and the control group in terms of the test score/omissions (both $F(1, 54) = 10.92$, $p < 0.005$) and the corrected test score ($F(1, 57) = 6.100$, $p < 0.05$). Participants in the experimental group with the same scores as participants in the control group would significantly outperform their counterparts. The performance results obtained are further visualized in Figure 2. Whereas the experimental group achieved an average test score of 12.23, a corrected test score of 9.23 and made only 2.77 omissions, the control group only achieved a test score of 10.29, a corrected test score of 7.71 and made 4.71 omissions. In this case, we did not run the second step of the analysis as for the strategy prompts since significant results of the first step suggest that an efficient threshold level has been chosen.

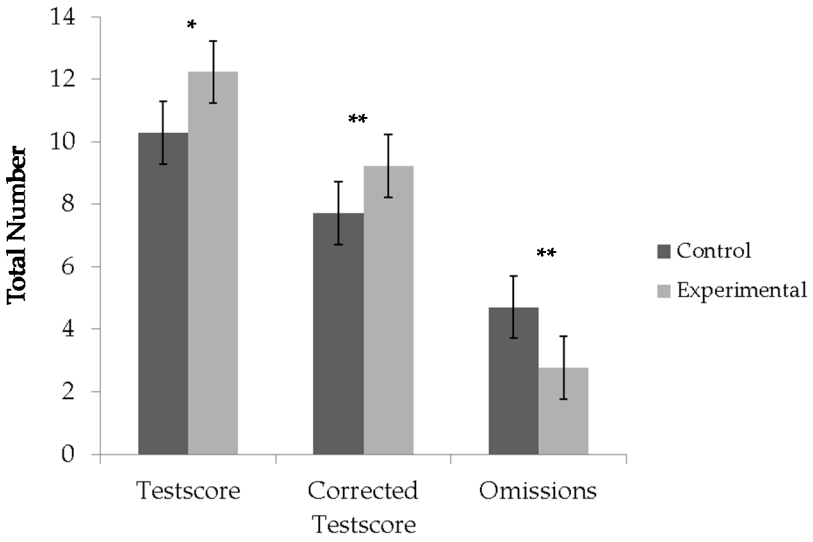


Figure 2. Mean differences between the control and experimental group. The bars represent the standard error. * $p < 0.01$, ** $p < 0.005$.

3.3.3. Explanation Prompt

In the first step of the analysis, participants of the control group were matched to those of the experimental group concerning their values for work effort. No significant performance differences for female participants with a value $\leq$15 and males with a value $\leq$18 were determined (thresholds determined for prompts; $p > 0.05$). No more efficient threshold level could be determined in step 2 of the analysis.

3.3.4. Verification Prompt

For work effort (threshold 20), we observed no significant difference in confidence between participants of the two groups ($p > 0.05$). For need for cognition (threshold 25), we observed the same non-significant trend. Please note, however, that judgment of learning, similar to need for cognition and work effort, correlated positively with performance outcomes. This does emphasize the importance of this form of self-assessment in online learning. In other words, there was no evidence that participants were overconfident. Indeed, they tended to be very accurate self-assessors of their learning progress. This means the verification prompt did not actually address overconfidence as overconfidence was not a concern with our participants. In response to this, we omitted the second step of the analysis as optimal thresholds are only required when overconfidence indeed exists—an issue for future research.

3.3.5. Motivation Prompt

No significant differences were found at the original threshold levels ($p > 0.05$). Subsequently we proceeded to the second step of the analysis to investigate the possibility of improving thresholds. Indeed, significant performance differences were found when we excluded participants who had scored exactly 12 on the computer self-efficacy scale. Whereas the control group reached an average test score of 10.68 and made 4.32 omissions, the experimental group actually performed significantly better and obtained a test score of 11.16 while making only 3.84 omissions. This suggests better performance for participants with low computer self-efficacy in the experimental group when prompted. The threshold analysis for the judgment of learning score revealed a similar pattern. Performance differences emerge below the threshold of 75 for judgment of learning. Whereas the control group reached a test score of 10.78 and made 4.77 omissions, the corresponding participants of the experimental group reached a higher test score of 12.12 and made fewer (only 2.88) omissions.

4. Discussion

The future of learning will require more and more engagement with online resources, such as e-modules. As a result, it becomes increasingly important that the online learning materials offer appropriate guidance to the learners—and recognize the unique aspects of an individual participant. Such guidance may come in numerous forms, including prompts that enable the learner to regulate all of those (cognitive, motivational, or behavioral) processes that are particular to the individual learner's approach [6]. The present work attempted to outline the merits of personalized support during e-learning, based on a variety of learner characteristics and the work around self-regulation.

Prompting as utilized in this study provides designers with one way in which they may be able to increase performance of those who may be more likely to underperform based on their characteristics. Different prompts may do so by either supporting recall or encouraging learners to become more cognizant of their actual performance in computer-supported learning systems [25]. Using existing personality and strategy evidence, we demonstrated one approach to personalization. Our work, therefore, builds on the previous research on the importance of considering different learning characteristics such as personality and learning strategies in online settings [1–3].

The focus of our study was to find answers to two research questions. Specifically, we asked if participants perform better when prompted compared to a participants with an identical learner profile when not prompted (RQ1). This was examined using matched samples from both the control group (not prompted) and the experimental group. We found that prompting (RQ1) was only effective when the assessment prompt is presented. This finding builds on the ideas around the testing effect [21], that by having learners recall what they have learned, they will be able to recall information better [22].

A general summary provides starting points for recommendations. First, confidence was the most robust predictor of performance as greater confidence correlated positively with performance. Good performance in previous assessments may raise future performance expectations. This means that collecting confidence reports may generate a helpful baseline measure on which to build subsequent prompts in other online systems. However, it may also be worthwhile considering just how robust and fragile confidence is in online assessment (especially because instantaneous feedback can be integrated), as understanding this may provide further insight into the relationship to performance; Second, computer self-efficacy correlated negatively with test performance—an effect that we would expect will diminish over time as learners become more accustomed to new technologies. As we obtained these findings with younger participants (a student sample) rather than older participants (who may not have used e-learning during their studies) suggests that computer self-efficacy may not be merely an element of age or experience, but potentially a measure of perceived capability when assessment is moved from traditional paper to online systems. Given the number of different systems users may encounter online, personalization of a system may need to account for the heterogeneous background of users.

The lack of support for the other prompts may be due to the thresholds selected. This question was tackled as well: In the case the prompts do not lead to better performance, can more efficient thresholds be determined in terms of performance outcome (RQ2)? Specifically, we considered if our prompting thresholds were effective and possibilities to improve these. From a strict statistical point of view our thresholds did not operate as effectively as planned. The strategy, explanation, and verification prompts did not perform satisfactorily, regardless of the threshold, which suggests they should either not be implemented in future studies or must be designed differently.

Limitations and Future Research

A number of potential limitations apply. First, many of the learner characteristics used for implementing prompts could be considered both: affective states, as well as more stable trait characteristics. Whether the variables of the learner profile should be regarded more a state or a trait when used in an e-learning environment, may be focus of future studies; Second, by employing a dichotomous prompting approach based on whether or not participants scored high on specific

learning characteristic reduces variance [43]. An alternative approach could be to simply use percentile, or even decentile, rankings and create corresponding prompts or use an item response theoretical approach (for successful applications in online learning settings see work by [35,44,45]); Third, our information about participants was collected using self-report and was limited to a small set of variables; Fourth, prompts were kept relatively short and it was assumed that they were self-explanatory. This may have disadvantaged some participants with poorer comprehension skills. In addition, it would be interesting examining additive or inhibitive effects between prompts. The latter was outside the remit of the current study but Kauffman and colleagues [15] may be a good starting point; And fifth, relying on null hypothesis testing [46] and adherence to *p*-values as the determinants of what may be considered meaningful has not only been criticized by leading statisticians [47] but may also be a suboptimal approach to optimize prompting. For example, Reisslein [48] explored different prompt formats and presentation effects. They interpreted non-significant results as an indicator that the prompts were effective across all groups where they were employed (rather than ineffective). Krause and Stark [49] also observed no significant prompt difference in performance when they asked students to engage in active problem-solving with or without reflection prompts. In their case, the descriptive statistics showed that purely numerically, performance was indeed higher for the prompted group, but not at $p < 0.05$ (they reported a difference $p = 0.11$).

Based on these limitations the authors would like to make five suggestions for future research. First, there was tentative evidence that the motivation prompt might be of use if thresholds are refined. This suggests that careful analysis of prompts may, even when prompts fail to generate the expected outcomes, provide starting points to optimize future implementations; Second the readability indices used to ensure that the e-modules are fairly comparable are based on linguistic features. More elaborated approaches consider cognitive features and processes relevant to text processing [50]. Using Coh-Metrix as a readability index, for example, may shed more light on the interplay between text difficulty, prompting and learner characteristics; Third, more work is needed to understand the testing effect and how different prompts may come into play. For example, future studies may investigate whether the assessment prompt has to be based on the learner profile or may work for any student independent of their learner profile; Fourth, future longitudinal employment of learning profiles, such as the one we utilized, may provide insight into the stability and effectiveness of prompt interventions [51]; Fifth, and finally, the challenges with the design of thresholds will be to identify optimal points where performance will be sufficiently enhanced, and this will require reiterative designs that improve on previous models. Sanabria and Killeen [52] suggested that replication statistics may be particularly helpful, and potentially more so than null hypothesis testing, in order to test replicated effects. Given the incremental work required to find optimal thresholds, we would support the additional use of these statistics in future work.

5. Conclusions

In line with previous work on personalized prompting, we tested a number of prompts which we introduced together with an overview of how we operationalized personalized prompting. These were based on a number of learning and personality characteristics predicted to influence learning. The use of a matched case-control design enabled us to compare the effects. Unfortunately, the motivation, the verification, the strategy and the explanation prompt did not improve performance. Since we examined different thresholds in an exploratory manner, it is likely that these prompts were simply not effective the way they were designed.

We observed a significant benefit of the assessment prompt, which is in line with the testing effect, which proved an excellent starting point for future studies. A particular research question may address the issue whether the assessment prompt has to be based on the learner profile or works independently. All learning characteristics were significant correlates of at least one of the performance variables. Additional thought was given to the role of confidence, computer self-efficacy, and opportunities for threshold optimization in the hope to provide further starting points for future work on personalization.

Acknowledgments: The authors gratefully acknowledge the support of Christian Stamov Roβnagel, Dirk Schulte am Hülse, as well as the various instructors and participants. In addition, we thank the editors and reviewers for their helpful comments and guidance. The project did not receive specific funding, but benefited from them previous work supervised by Christian Stamov Roßnagel and sponsored through a grant (Grant No. 01PF07044) from the German Ministry of Education and Research (BMBF) in 2011 to 2013.

Author Contributions: Joy Backhaus and Debora Jeske designed the study and implemented the data collection effort. Herbert Poinstingl and Sarah Koenig provided statistical support and institutional support during the write-up and preparation of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, C.-M. Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Syst. Appl.* **2009**, *36*, 8816–8829. [CrossRef]
2. Köck, M.; Paramythis, A. Activity sequence modelling and dynamic clustering for personalized e-learning. *User Model. User-Adapt. Interact.* **2011**, *21*, 51–97. [CrossRef]
3. Özpolat, E.; Akar, G.B. Automatic detection of learning styles for an e-learning system. *Comput. Educ.* **2009**, *53*, 355–367. [CrossRef]
4. Chen, N.-S.; Wei, C.-W.; Wu, K.-T.; Uden, L. Effects of high level prompts and peer assessment on online learners' reflection levels. *Comput. Educ.* **2009**, *52*, 283–291. [CrossRef]
5. Demetriadis, S.N.; Papadopoulos, P.M.; Stamelos, I.G.; Fischer, F. The effect of scaffolding students' context-generating cognitive activity in technology-enhanced case-based learning. *Comput. Educ.* **2008**, *51*, 939–954. [CrossRef]
6. Bannert, M.; Reimann, P. Supporting self-regulated hypermedia learning through prompts. *Instr. Sci.* **2012**, *40*, 193–211. [CrossRef]
7. Dunnette, M.D.; Hough, L.M. *Handbook of Industrial and Organizational Psychology*; Consulting Psychologists Press: Palo Alto, CA, USA, 1991; Volume 2.
8. Cacioppo, J.T.; Petty, R.E.; Feinstein, J.A.; Jarvis, W.B.G. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychol. Bull.* **1996**, *119*, 197–253. [CrossRef]
9. Biggs, J.; Kember, D.; Leung, D.Y. The revised two-factor study process questionnaire: R-spq-2f. *Br. J. Educ. Psychol.* **2001**, *71*, 133–149. [CrossRef] [PubMed]
10. Hua, T.; Williams, S.; Hoi, P. Using the Biggs' Study Process Questionnaire (SPQ) As a Diagnostic Tool to Identify "At Risk" Students—A Preliminary Study. Available online: http://static1.1.sqspcdn.com/static/f/1751776/20693261/1350728726863/Identifying-at-risk-students-with-spq.pdf?token=H8CwQ1NiGyJ5iHJFk34ah5weFlQ%3D (accessed on 9 February 2017).
11. Compeau, D.R.; Higgins, C.A. Computer self-efficacy: Development of a measure and initial test. *MIS Q.* **1995**, *19*, 189–211. [CrossRef]
12. Rawson, K.A.; Dunlosky, J.; McDonald, S.L. Influences of metamemory on performance predictions for text. *Q. J. Exp. Psychol. Sect. A* **2002**, *55*, 505–524. [CrossRef] [PubMed]
13. Maki, R.H.; Shields, M.; Wheeler, A.E.; Zacchilli, T.L. Individual differences in absolute and relative metacomprehension accuracy. *J. Educ. Psychol.* **2005**, *97*, 723–731. [CrossRef]
14. Chi, M.T.; De Leeuw, N.; Chiu, M.-H.; LaVancher, C. Eliciting self-explanations improves understanding. *Cogn. Sci.* **1994**, *18*, 439–477.
15. Kauffman, D.F.; Ge, X.; Xie, K.; Chen, C.-H. Prompting in web-based environments: Supporting self-monitoring and problem solving skills in college students. *J. Educ. Comput. Res.* **2008**, *38*, 115–137. [CrossRef]
16. Berthold, K.; Eysink, T.H.; Renkl, A. Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instr. Sci.* **2009**, *37*, 345–363. [CrossRef]
17. Pintrich, P.R. *The Role of Goal Orientation in Self-Regulated Learning*; Academic Press: San Diego, CA, USA, 2000.
18. Devolder, A.; van Braak, J.; Tondeur, J. Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *J. Comput. Assist. Learn.* **2012**, *28*, 557–573. [CrossRef]

19. Moreno, R.; Mayer, R.E.; Spires, H.A.; Lester, J.C. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cogn. Instr.* **2001**, *19*, 177–213. [CrossRef]
20. Sitzmann, T.; Ely, K. A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychol. Bull.* **2011**, *137*, 421. [CrossRef] [PubMed]
21. Nungester, R.J.; Duchastel, P.C. Testing versus review: Effects on retention. *J. Educ. Psychol.* **1982**, *74*, 18–22. [CrossRef]
22. Roediger, H.L.; Karpicke, J.D. Test-enhanced learning taking memory tests improves long-term retention. *Psychol. Sci.* **2006**, *17*, 249–255. [CrossRef] [PubMed]
23. Schunk, D.H.; Zimmerman, B.J. *Self-Regulated Learning: From Teaching to Self-Reflective Practice*; Guilford Press: New York, NY, USA, 1998.
24. Thiede, K.W.; Dunlosky, J.; Griffin, T.D.; Wiley, J. Understanding the delayed-keyword effect on metacomprehension accuracy. *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31*, 1267–1280. [CrossRef] [PubMed]
25. Schworm, S.; Renkl, A. Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Comput. Educ.* **2006**, *46*, 426–445. [CrossRef]
26. Hacker, D.J.; Bol, L.; Horgan, D.D.; Rakow, E.A. Test prediction and performance in a classroom context. *J. Educ. Psychol.* **2000**, *92*, 160–170. [CrossRef]
27. Lauterman, T.; Ackerman, R. Overcoming screen inferiority in learning and calibration. *Comput. Hum. Behav.* **2014**, *35*, 455–463. [CrossRef]
28. Dunning, D.; Meyerowitz, J.A.; Holzberg, A.D. Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *J. Personal. Soc. Psychol.* **1989**, *57*, 1082–1090. [CrossRef]
29. Van Merriënboer, J.J. *Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training*; Educational Technology: Englewood Cliffs, NJ, USA, 1997.
30. Andersen, E.B. A goodness of fit test for the rasch model. *Psychometrika* **1973**, *38*, 123–140. [CrossRef]
31. Senter, R.; Smith, E.A. Automated Readability Index. *AMRL-TR. Aerosp. Med. Res. Lab.* **1967**, 1–14.
32. Flesch, R. A new readability yardstick. *J. Appl. Psychol.* **1948**, *32*, 221–233. [CrossRef] [PubMed]
33. Klare, G.R. Assessing readability. *Read. Res. Q.* **1974**, *10*, 62–102. [CrossRef]
34. De Cooman, R.; De Gieter, S.; Pepermans, R.; Jegers, M.; Van Acker, F. Development and validation of the work effort scale. *Eur. J. Psychol. Assess.* **2009**, *25*, 266–273. [CrossRef]
35. Barbeite, F.G.; Weiss, E.M. Computer self-efficacy and anxiety scales for an internet sample: Testing measurement equivalence of existing measures and development of new scales. *Comput. Hum. Behav.* **2004**, *20*, 1–15. [CrossRef]
36. Petty, R.E.; Cacioppo, J.T.; Kao, C.F. The efficient assessment of need for cognition. *J. Personal. Assess.* **1984**, *48*, 306–307.
37. Kember, D.; Gow, L. Cultural specificity of approaches to study. *Br. J. Educ. Psychol.* **1990**, *60*, 356–363. [CrossRef]
38. Jeske, D.; Backhaus, J.; Roßnagel, C.S. Evaluation and revision of the study preference questionnaire: Creating a user-friendly tool for nontraditional learners and learning environments. *Learn. Individ. Differ.* **2014**, *30*, 133–139. [CrossRef]
39. Jeske, D.; Backhaus, J.; Stamov Roßnagel, C. Self-regulation during e-learning: Using behavioural evidence from navigation log files. *J. Comput. Assist. Learn.* **2014**, *30*, 272–284. [CrossRef]
40. Jeske, D.; Stamov Roßnagel, C.; Backhaus, J. Learner characteristics predict performance and confidence in e-learning: An analysis of user behaviour and self-evaluation. *J. Interact. Learn. Res. (JILR)* **2014**, *25*, 509–529.
41. Tanaka, J.; Panter, A.T.; Winborne, W.C. Dimensions of the need for cognition: Subscales and gender differences. *Multivar. Behav. Res.* **1988**, *23*, 35–50. [CrossRef] [PubMed]
42. Martin, M.A. Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties. *Comput. Stat. Data Anal.* **2007**, *51*, 6321–6342. [CrossRef]
43. Cohen, J. The Cost of Dichotomization. *Appl. Psychol. Meas.* **1983**, *7*, 249–253. [CrossRef]
44. Daomin, X.; Mingchui, D. Appropriate learning resource recommendation in intelligent web-based educational system. In Proceedings of the 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, Zhangjiajie, China, 6–7 November 2013; pp. 169–173.
45. Stevens, R.; Beal, C.R.; Sprang, M. Assessing students' problem solving ability and cognitive regulation with learning trajectories. In *International Handbook of Metacognition and Learning Technologies*; Springer: New York, NY, USA, 2013; pp. 409–423.

46. Robinson, D.H.; Wainer, H. On the past and future of null hypothesis significance testing. *J. Wild. Manag.* **2002**, *66*, 263–271. [CrossRef]
47. Gigerenzer, G.; Krauss, S.; Vitouch, O. The null ritual. In *The Sage Handbook of Quantitative Methodology for the Social Sciences*; Sage Publications: Thousand Oaks, CA, USA, 2004; pp. 391–408.
48. Reisslein, J.; Atkinson, R.K.; Seeling, P.; Reisslein, M. Investigating the presentation and format of instructional prompts in an electrical circuit analysis computer-based learning environment. *IEEE Trans. Educ.* **2005**, *48*, 531–539. [CrossRef]
49. Krause, U.-M.; Stark, R. Reflection in example-and problem-based learning: Effects of reflection prompts, feedback and cooperative learning. *Eval. Res. Educ.* **2010**, *23*, 255–272. [CrossRef]
50. Abidin, M.J.Z.; Rezaee, A.A.; Abdullah, H.N.; Singh, K.K.B. Learning styles and overall academic achievement in a specific educational system. *Int. J. Humanit. Soc. Sci.* **2011**, *1*, 143–152.
51. Sitzmann, T.; Ely, K. Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. *J. Appl. Psychol.* **2010**, *95*, 132–144. [CrossRef] [PubMed]
52. Sanabria, F.; Killeen, P.R. Better statistics for better decisions: Rejecting null hypotheses statistical tests in favor of replication statistics. *Psychol. Sch.* **2007**, *44*, 471–481. [CrossRef] [PubMed]



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).