

LEARNING TO USE A DIGITAL WORKBENCH TO DEVELOP COMPETENCE-BASED TESTS IN VOCATIONAL EDUCATION: GUIDED OR EXPLORATIVE?

Josef Buchner, Anne-Cathrin Vonarx, Peter Pfänder and Michael Kerres

*Learning Lab, University of Duisburg-Essen
Universitätstr. 2, 45141 Essen, Germany*

ABSTRACT

This study arises from the question of how learners experience different learning arrangements while learning to develop competence-based tasks with a new digital workbench. The study compares a guided arrangement with worked examples to an exploratory learning scenario. Leading questions for the design of the study and analysis are as follows: Does the guided group develop more tasks? Does the explorative group test more functionalities of the digital workbench? Does the guided group report lower cognitive load and a more positive attitude toward the workbench?

The analysis of the elements created within the workbench reveals that the guided group was able to develop more tasks, while there was no difference in quality. The data show that the explorative group was able to test more functionalities; however, the difference was not significant. The results of the questionnaire show that the overall cognitive load also did not differ significantly. There are differences in the scales, which are discussed more deeply in this article. Regarding the further use of the digital workbench¹, both groups report a positive attitude. It can be concluded that learning to use a new digital environment for developing examination tasks should be supported; however, the form of guidance can be subject to further investigation.

KEYWORDS

Worked Examples, Cognitive Load, Role of Attitudes, Exploratory Learning, Scaffolded Learning, Task-Based Learning

1. INTRODUCTION

Which learning arrangement helps learners most? This question has occupied educational and psychological science for a very long time and has led to intensive discussions between researchers around the world. One of the best-known debates is between Kirschner et al. (2006) and Hmelo-Silver et al. (2007). While the former argue against unguided exploration (i.e., pure discovery learning), Hmelo-Silver et al. claim that discovery, exploratory and problem-based learning are effective under certain conditions, such as through scaffolding. More recent studies report that different instructional approaches also pursue different learning goals; therefore, different learning goals need different instructional methods (Koedinger et al., 2012; Renkl, 2015). For example, the promotion of interest, motivation or attitude toward scientific disciplines in explorative learning scenarios can be an important (learning) goal (Newman and DeCaro, 2019). Likewise, different approaches can also moderate the learning process (Hmelo-Silver, 2004), as demonstrated in Likourezos and Kalyuga (2017). Their study provided evidence that participants in an unguided group accepted the learning of new content as a challenge more strongly than a guided group. The authors explain that this compensated for the higher cognitive load that was otherwise associated with this type of learning approach (Likourezos and Kalyuga, 2017, p. 214).

According to cognitive load theory (CLT), excessive cognitive load can unnecessarily strain working memory and hinder learning. This is especially relevant for beginners who have little prior knowledge of a certain domain (Sweller, 1988; Sweller et al., 2019). If they are confronted with low-guidance learning

¹ Project ASPE: <https://learninglab.uni-due.de/forschung/projekte/aspe-kompetenzorientierte-pr-fungen>

methods, these learners are quickly overwhelmed and must invest more time, e.g., to solve a task or a problem (Sweller et al., 2007, p. 116). One way to counteract this challenge is to reduce extraneous cognitive load. CLT has identified many effects related to this phenomenon, such as the worked example effect (Sweller, 2020). This effect has been demonstrated in many empirical studies and shows that the use of worked examples to learn domain-specific knowledge is more effective than teaching approaches without such guidance. The reduction of extraneous load also leads to better results in understanding and transferring knowledge to other problems (Renkl, 2014; van Gog et al., 2011; Wittwer and Renkl, 2010).

Worked examples are elaborated samples of problems that lead step by step to the solution to the problem. Novices can use them to follow an expert's approach, and as their level of expertise increases, they can detach themselves from the supportive materials. Worked examples mostly use practical, highly structured problems (Atkinson and Renkl, 2007). However, there are also examples of worked examples for less structured problems, e.g., writing, argumentation training or studying teaching methods (van Gog and Rummel, 2010). The learning of digital tools can also be supported with the help of worked examples (Hilbert and Renkl, 2009).

It must be noted that the worked examples effect is sufficient for beginners. Experts or advanced learners may perceive worked examples as not conducive to learning. This effect is called the expertise reversal effect and has been proven in many empirical studies (Kalyuga, 2007).

Furthermore, worked examples limit the exploration of other possible solutions, which is known as learning through failure (Kapur, 2008). Achieved learning goals within a worked example learning scenario are usually based on a very clear and restrictive structure, which can hinder the acquisition of learning objectives beyond the constructed setting (Hmelo-Silver, 2004).

2. PRESENT STUDY

Based on the abovementioned findings on learning with worked examples, this study examines how a fully guided instructional design with worked examples affects different learning goals compared to an explorative approach. The first learning objective involves participants' creation of tasks for vocational training examination with a new digital platform, called the digital workbench. This objective is divided into two parts: the quantity and the quality of the tasks created. The second learning objective is the examination of the functions of the digital workbench. The cognitive load as well as the attitude of the participants toward the workbench will be examined since learner characteristics such as attitude can influence the perception of cognitive load (Scheiter et al., 2009). The following hypotheses will be tested:

- H1: The group guided through worked examples scores better than the unguided group on the 1st learning objective (quantity and quality of the prepared examination tasks).
- H2: The unguided group explores more functionalities (learning objective 2) of the workbench than the guided group.
- H3: The guided group reports a lower cognitive load than the unguided group.
- H4: The guided group shows more positive attitudes toward the workbench than the unguided group.
- H5: Attitude toward the workbench positively influences the perceived cognitive load.

3. METHOD

3.1 Design and Participants

A total of 31 participants (four women, 22 men, five did not specify) with an average age of 50.6 (SD = 8.9) years participated in the study. The participants worked as teachers and research assistants for schools and a state organization for nationwide examinations in vocational education and training. We used a randomized experimental design. Sixteen participants were randomly assigned to the fully guided worked examples group (FG) and 15 to the unguided exploratory group (UG).

3.2 Materials

3.2.1 Prototype of the Workbench

Within the framework of the ASPE project², a digital workbench was developed to support the participants of this study in the future creation of examination tasks. In this study, the first prototype was tested for the first time. An examination in vocational education and training consists of several parts, all of which can be created in the digital workbench and then combined to form an examination. The prototype allows the research team to access all the content created. This created content is evaluated and used to determine how successfully learning objectives 1 and 2 have been achieved.

3.2.2 Learning Tasks and Objectives

The tasks were handed out to the participants in written paper-based form. Each element of an exam had to be worked out at least once by the participants in the same way as for a real exam. The other learning objective was to explore the functionalities of the digital workbench.

3.2.3 Instructional Worked-Out Example

The step-by-step instructions were issued only to the guided group in paper-based form. The manual consisted of text-image combinations with regard to design principles of the cognitive theory of multimedia learning (CTML) (Mayer, 2019).

3.2.4 Measurement of Cognitive Load

Cognitive load was measured with the Nasa Task Load Index (NASA-TLX). The NASA-TLX is a self-reported questionnaire with six subscales on a scale of 0-20 (Hart, 2006; Hart and Staveland, 1988). It is used frequently because it meets the requirement of multidimensional measurement of cognitive load compared to one-dimensional scales (de Jong, 2010). Participants filled out the NASA-TLX after working with the workbench (Cronbach's alpha = 0.7).

3.2.5 Survey of Attitudes

We used an adapted version of the personal digital assistants' attitude survey (Cheng, 2017) to investigate the participants' attitudes toward the use of the workbench. This questionnaire consists of four scales: usability ($\alpha = 0.9$, four items), usefulness ($\alpha = 0.9$, four items), concerns ($\alpha = 0.6$, three items) and intention to use ($\alpha = 0.7$, three items). The Likert scale ranges from 1 (do not agree at all) to 5 (agree completely).

3.2.6 Procedure of Measurement of Cognitive Load and Attitude

The participants were randomly assigned to one of the two conditions, guided with worked examples (FG group) or explorative without guidance (UG group), by means of colored markings on the name tag. Two to three participants met in a small group to test the workbench. Each FG group received the worked example and the paper-based task. The small groups in the UG group received only the printout with the task. After the intervention, the participants completed the questionnaires on cognitive load and attitudes. All elements created in the workbench were saved anonymously.

4. RESULTS

For the evaluation of the learning success regarding the two learning objectives, the data from the workbench were first prepared in tabular form, and then the values were calculated together with the values from the questionnaires on cognitive load and attitude in SPSS 26.

² <https://learninglab.uni-due.de/forschung/projekte/aspe-kompetenzorientierte-pr-fungen>

4.1 Descriptive Statistics

In Table 1, we present the descriptive statistics for each variable. Learning success is divided into learning objective 1 with the quantitative and qualitative components. Learning objective 2 involves testing the functions of the workbench.

Table 1. Descriptive statistics with means and standard deviations for all variables

Variable	FG (n = 16)		UG (n = 15)	
	M	SD	M	SD
Learning Obj. 1 – quantitative	3.50	1.38	1.71	0.95
Learning Obj. 1 – qualitative	6.50	5.68	4.00	3.00
Learning Obj. 2 – functions	8.00	5.40	10.00	6.90
Cognitive Load				
Mental Demand	10.06	5.13	10.53	5.84
Physical Demand	5.12	6.12	4.40	3.23
Temporal Demand	4.00	3.76	8.92	4.73
Performance*	8.50	4.65	12.20	4.78
Effort	8.00	5.48	8.33	4.98
Frustration	6.25	5.62	8.80	5.67
Overall CL	6.99	3.31	8.85	3.05
Attitudes				
Usability	3.14	0.46	2.98	0.50
Usefulness	3.09	1.11	3.68	1.16
Concerns	1.46	0.47	1.58	0.56
Intention to Use	3.65	0.93	3.98	0.92
Overall Attitude	2.83	0.44	3.00	0.46

*Note: Scale performance is reverse coded.

4.2 H1: Does the Guided Group Score Better?

To test the first hypothesis, “the group with the support of worked examples scores better than the unguided group on the 1st learning objective (quantity and quality of the prepared examination tasks”, we calculated an independent t-test with the values for small groups per condition (Table 2). Levene's test for equality of variance showed no significant values for learning success quantitatively ($p = 0.07$) or qualitatively ($p = 0.46$), so we can assume homogeneity of variance.

Table 2. Results of the independent t-test to test hypothesis one

Variable	FG		UG		t(11)	p	Cohen's d
	M	SD	M	SD			
Learning Obj. 1 – quantitative	3.50	1.38	1.71	0.95	-2.75	0.019	1.50
Learning Obj. 1 – qualitative	6.50	5.68	4.00	3.00	-1.02	0.33	-

The results in Table 2 support part of our first hypothesis. The FG group differed significantly from the UG group in quantitative learning success, $t(11) = -2.76$, $p = 0.019$. This means that the FG group solved the required task significantly more often than the UG control group. The effect size according to Cohen (1992) was high ($d = 1.50$). No significant differences were found in the quality of the actual tasks for final vocational examinations that were created by the participants. The FG group created more tasks but did not create them at a higher quality than the lesser tasks created by the UG group ($t(11) = -1.02$, $p = 0.33$).

4.3 H2: Does the Explorative (Unguided) Group Test More Functionalities?

To answer hypothesis two, a t-test was again calculated for independent samples (Table 3), this time with the variable functions explored as a dependent variable. Levene's test again showed no significance. We again assume homogeneity of variance ($p = 0.41$).

Table 3. Results of the independent t-test to test hypothesis two

Variable	FG		UG		t(11)	p	Cohen's d
	M	SD	M	SD			
Learning Obj. 2 – functions	8.00	5.40	10.00	6.90	0.57	0.58	-

The mean values in Table 3 show that the UG group explored the functionality of the workbench slightly more. However, this difference is not significant, $t(11) = 0.57$, $p = 0.58$, so we reject this hypothesis.

4.4 H3: Does the Guided Group Report a Lower Cognitive Load?

To determine whether the unguided group (UG) reported a higher cognitive load, we compared the mean values of each scale and the overall cognitive load using a t-test. Except for the physical demand scale (Levene's test, $p = 0.029$), all scales showed variance homogeneity (p -value between 0.45 - 0.90). For the overall cognitive load, there was no significant effect between the groups, $t(29) = 1.63$, $p = 0.12$. Additionally, the scales of mental demand ($p = 0.81$), physical demand ($p = 0.67$), effort ($p = 0.86$) and frustration ($p = 0.22$) did not differ significantly between the FG group and the UG group.

We found significant differences for the temporal demand and performance scales (Table 4).

Table 4. Significant results of the independent t-test to test hypothesis three on cognitive load

Variable	FG		UG		t(29)	p	Cohen's d
	M	SD	M	SD			
Temporal Demand	4.00	3.76	8.92	4.73	3.12	0.004	1.21
Performance*	8.50	4.65	12.20	4.78	2.18	0.037	0.78

*Note: Scale performance is reverse coded.

As Table 4 shows, the participants of the UG group found the processing of the task in the workbench more time-consuming than did the FG group, and the effect size according to Cohen (1992) was large, $d = 1.21$.

We also see a large effect in the results of the participants' own assessment of the performance that can be achieved with regard to the task ($d = 0.78$). This means that the group with the worked examples believed more strongly that they could successfully complete the task set compared to the UG group.

4.5 H4: Does the Guided Group Show More Positive Attitudes Toward the Workbench?

We must reject this assumption. The results from the independent t-test do not show significant differences between the FG and UG groups for any scale or for the summary of all scales (overall attitude). We interpret the results as indicating that both free exploration and a guided instructional approach fostered positive attitudes toward our developed workbench.

4.6 H5: Does the attitude toward the workbench influence the perceived cognitive load?

To test hypothesis 5, we calculated linear regression models with the overall cognitive load as the dependent variable and the individual attitude scales as independent variables. The results are shown in Table 5.

Table 5. Effect of attitude on cognitive load

Variable	Unstandardized		Standardized		Coefficients Std. Error	
	FG	UG	FG	UG	FG	UG
(Constant)	19.19***	6.45				
Usability	-3.92**	-1.23	-0.54**	-0.20	1.32	1.78
Usefulness	-1.43*	1.50	-0.48*	0.57	0.66	1.07
Concerns	0.52	3.00	0.07	0.54	1.38	1.89
Intention to Use	1.03	-1.02	0.29	-0.31	0.72	1.28
R ²	0.67	0.36				
corr. R ²	0.55	0.11				
F-FG(4,11), F-UG(4,10)	5.63**	1.41				
f ²	1.11	-				

*p<0.05, **p<0.05, ***p<0.01

As we can see in Table 5, the model is significant only for the FG group ($F-FG(4,11) = 5.63, p = 0.01, N = 16$). The effect size $f^2 = 1.11$ corresponds to a strong effect according to Cohen (1992). The attitude toward the workbench among the participants of the FG group therefore had an influence on the cognitive load. Only the scales usability ($p = 0.01$) and usefulness ($p = 0.05$) were significant. If the positive attitude toward these two variables decreases, cognitive load can increase. This effect is stronger for usability (unstandardized beta = -3.92) than for usefulness (unstandardized beta = -1.43). In the UG group, however, we did not find any significant influences for any of the attitude scale variables that can be reported as predictors of cognitive load ($F-UG(4,10) = 1.41, p = 0.30, N = 15$). We also calculated a linear regression with the overall attitude scale as an independent variable for both groups. The model was not significant for the FG group, $F-FG(1,14) = 1.70, p = 0.21, N = 16$, or for the UG group, $F-UG(1,13) = 0.78, p = 0.39, N = 15$.

Based on the results found in 4.3, we calculated linear regression models with temporal demand and performance as dependent variables.

A scale from the attitudes scale was not a significant predictor for the sense of temporal demand in either group; $F-FG(4,11) = 1.49, p = 0.27$ and $F-UG(4,8) = 1.72, p = 0.24$.

For the variable performance, it was found that usability ($p = 0.02$) and usefulness ($p = 0.012$) were significant predictors in the FG group; $F-FG(4,11) = 5.38, p = 0.012, f^2 = 1.08$. This was not true for the UG group. None of the variables showed significant values in the sense of assumed performance; $F-UG(4,10) = 2.07, p = 0.16$.

5. DISCUSSION

This study examined how different learning arrangements affect learning goals, cognitive load, and attitudes toward a new digital workbench. One learning objective was to create vocational education examination tasks using the workbench, and quantitative and qualitative aspects were investigated. In terms of quantity, it was found that the group led by worked examples (FG) was more successful than the explorative group (UG). No differences were found in regard to the quality of the tasks created by the participants. Both groups showed similar performances. The second learning objective was to explore the new workbench and test its functions. Here, the descriptive statistics data show a small advantage in favor of the UG group; however, the difference is not significant. These results are consistent with other research results on worked examples. Highly structured and clear tasks are much better fulfilled by guided groups (Atkinson and Renkl, 2007). This was not the case for more explorative aspects of the task, where both groups performed similarly.

The overall cognitive load did not differ significantly between the groups, and only the scales of temporal demand and performance differed significantly. The UG group found working on the tasks more time consuming than the FG group. The FG group showed much stronger confidence in their own performance and reported higher values for performance. The time requirement for unguided instructional designs was also reported by Kirschner et al. (2006). Instructional designers should take this into account when designing learning arrangements and, in the case of more explorative approaches, give learners more time to solve tasks. Perceived performance can have a lasting effect on learning success, as Bandura's theory on self-efficacy shows (Bandura, 1977). Instructional designers should also be aware of the extent to which explorative learning environments may hinder learners' confidence in their success. No differences were found regarding attitudes toward the workbench. Both groups reported high values on all scales and were positive about further use.

The scales of usability and usefulness were found to be predictors of cognitive load, but only in the FG Group. For learning arrangements that are digitally supported, it can be concluded that the handling of the respective platform or technology should be perceived as simple by the users. With this in mind, the potential risk of cognitive load can be minimized. In conclusion, the technology used should also help to process certain tasks; otherwise, the risk of cognitive overload increases.

ACKNOWLEDGEMENT

The ASPE project is funded by the Federal Ministry of Education and Research, funding code: 21AP007AA.

REFERENCES

- Atkinson, R.K., Renkl, A., 2007. Interactive Example-Based Learning Environments: Using Interactive Elements to Encourage Effective Processing of Worked Examples. *Educational Psychology Review* 19, 375–386. <https://doi.org/10.1007/s10648-007-9055-2>
- Bandura, A., 1977. Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review* 84, 191–215.
- Cheng, K.-H., 2017. Reading an augmented reality book: An exploration of learners' cognitive load, motivation, and attitudes. *Australasian Journal of Educational Technology* 33, 53–69. <https://doi.org/10.14742/ajet.2820>
- Cohen, J., 1992. A Power Primer. *Quantitative Methods in Psychology* 112, 155–159.
- de Jong, T., 2010. Cognitive load theory, educational research, and instructional design: some food for thought. *Instr Sci* 38, 105–134. <https://doi.org/10.1007/s11251-009-9110-0>
- Hart, S.G., 2006. NASA-Task Load Index (NASA-TLX); 20 Years Later., in: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Presented at the Human Factors and Ergonomics Society 50th Annual Meeting, HFES, Santa Monica, pp. 904–908.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Hilbert, T.S., Renkl, A., 2009. Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior* 25, 267–274. <https://doi.org/10.1016/j.chb.2008.12.006>
- Hmelo-Silver, C.E., 2004. Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review* 16, 235–266. <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3>
- Hmelo-Silver, C.E., Duncan, R.G., Chinn, C.A., 2007. Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist* 42, 99–107. <https://doi.org/10.1080/00461520701263368>
- Kalyuga, S., 2007. Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educ Psychol Rev* 19, 509–539. <https://doi.org/10.1007/s10648-007-9054-3>
- Kapur, M., 2008. Productive Failure. *Cognition and Instruction* 26, 307–313. <https://doi.org/10.1080/07370000802212669>
- Kirschner, P.A., Sweller, J., Clark, R.E., 2006. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist* 41, 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Koedinger, K.R., Corbett, A.T., Perfetti, C., 2012. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science* 36, 757–798. <https://doi.org/10.1111/j.1551-6709.2012.01245.x>

- Likourezos, V., Kalyuga, S., 2017. Instruction-first and problem-solving-first approaches: alternative pathways to learning complex tasks. *Instructional Science* 45, 195–219. <https://doi.org/10.1007/s11251-016-9399-4>
- Mayer, R.E., 2019. Thirty years of research on online learning. *Applied Cognitive Psychology* 33, 152–159. <https://doi.org/10.1002/acp.3482>
- Newman, P.M., DeCaro, M.S., 2019. Learning by exploring: How much guidance is optimal? *Learning and Instruction* 62, 49–63. <https://doi.org/10.1016/j.learninstruc.2019.05.005>
- Renkl, A., 2015. Different roads lead to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice* 1, 79–90. <https://doi.org/10.1080/23735082.2015.994255>
- Renkl, A., 2014. Toward an Instructionally Oriented Theory of Example-Based Learning. *Cognitive Science* 38, 1–37. <https://doi.org/10.1111/cogs.12086>
- Scheiter, K., Gerjets, P., Vollmann, B., Catrambone, R., 2009. The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction* 19, 387–401. <https://doi.org/10.1016/j.learninstruc.2009.02.004>
- Sweller, J., 2020. Cognitive load theory and educational technology. *Education Tech Research Dev* 68, 1–16. <https://doi.org/10.1007/s11423-019-09701-3>
- Sweller, J., 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 257–285.
- Sweller, J., Kirschner, P.A., Clark, R.E., 2007. Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries. *Educational Psychologist* 42, 115–121. <https://doi.org/10.1080/00461520701263426>
- Sweller, J., van Merriënboer, J., Paas, F.G.W.C., 2019. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ Psychol Rev* 31, 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- van Gog, T., Kester, L., Paas, F., 2011. Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology* 36, 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>
- van Gog, T., Rummel, N., 2010. Example-Based Learning: Integrating Cognitive and Social-Cognitive Research Perspectives. *Educational Psychology Review* 22, 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- Wittwer, J., Renkl, A., 2010. How Effective are Instructional Explanations in Example-Based Learning? A Meta-Analytic Review. *Educational Psychology Review* 22, 393–409. <https://doi.org/10.1007/s10648-010-9136-5>