REPLICATING A STUDY WITH TASKS ASSOCIATED WITH THE EQUALS SIGN IN AN ONLINE ENVIRONMENT

Morten Elkjær[1] & Jeremy Hodgen[2]

[1]Danish School of Education (DPU) Aarhus University,

[2]University College London (UCL)

This paper presents a case study of a conceptual replication study. We replicated the famous and widely cited task presented in Falkner et al. (1999), 8+4=__+5. In contrast to the original study, we administered the task with the same age group (Grade 6) in a different system (Denmark) and via a large-scale online learning environment (OLE), with a larger sample and two decades later. Our replication indicates that the Danish students performed very significantly better than the students in the original study. We discuss why this is the case and argue that OLEs such as the one we used provide an important opportunity to replicate, and thus better understand, similar results.

INTRODUCTION

There is an increasing interest in replication studies in mathematics education at PME (e.g. Inglis et al., 2018) and beyond (e.g. Jankvist et al., 2021). This interest stems from the replication crisis in psychology research, which has highlighted a large proportion of false-positive results (e.g., Open Science Collaboration, 2015). In part, this may be due to the high degree of flexibility in quantitative and experimental researchers' analytic and design choices (Simmons et al., 2011). The imperative for replication studies in mathematics education is, however, broader than this. Aguilar's (2020) literature review highlights the majority of studies published even in respected mathematics education journals are small-scale and hence influenced by contextual factors that are poorly understood. Hence, replication can perform a crucial function in deepening and extending the validity of findings, because "[t]hrough variations to studies, we can delineate the bounds of the original study's findings" (Melhuish & Thanheiser, 2018, p. 106). Jankvist et al. (2021) emphasises that replication studies are important in the mathematics education community because they enable a more deeply understanding of the phenomena and results. Replication studies can help clarify under which conditions a particular finding is true or not and replication whether the results are stable over time, across different educational systems or different populations (e.g. Cai et al., 2018). Aguilar (2020) concludes that knowing more about the conditions that make it possible for a research finding to take place, and the boundaries of where it remains true, advances our research field as it allows us to broaden our understanding of the contextual variables under which the research finding occurs. This in turn has direct implications for the implementation of research findings in practice.

2022. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.), *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 251-258). PME.

In this paper, we present the data of a conceptual replication study (Aguilar, 2020) of the study presented in Falkner et al. (1999) and Carpenter et al. (2003), reporting findings from the use of their famous task $8 + 4 = \_ + 5$. This result that is widely cited in the literature on equivalence (e.g. Knuth et al., 2006). From the study presented in Falkner et al. (1999) we learn that an entire range of 145 sixth grade students provided 12 and/or 17 as the number that should go in the empty space. Students argue that 12 is the answer, because the numbers on the left together makes 12, neglecting the meaning of the +5 on the right side and reflecting an operational, rather than a relational, understanding of the equal sign (Knuth et al., 2006). Others argue that what goes in the empty space is the value of all the numbers added resulting in 17. In the two original studies, we are presented with the following data;

Percent of children offering various solutions to 8 + 4 = □ + 5

	Answers Given					Number of
Grade	7	12	17	12 and 17	Other	Children
1	0	79	7	0	14	42
1 and 2	6	54	20	0	20	84
2	6	55	10	14	15	174
3	10	60	20	5	5	208
4	7	9	44	30	11	57
5	7	48	45	0	0	42
6	0	84	14	2	0	145

Table 1: Data from answer provided to $8 + 4 = \_\_ + 5$ (Falkner et al., 1999, p. 223)

We have recreated the above task with two additional variations $4 + \_ = 7 + 5$ and $6 + \_ = 4 + 5$, and implemented them in a Danish OLE called *matematikfessor.dk*. The variations are made in order to investigate the bounds of Falkner et al.'s (1999) findings. The first variation uses the same format as the original task but the empty space has been moved to the left side of the equals sign. This is done in order to investigate how willing students are to put the number 3, completing the sum $4 + 3 = 7$, ignoring the number 5 at the end, similarly to the original task. We did however not expect the students to be willing to put in 16 (the total sum of the numbers present) but were curious whether the students would put 12 completing the sum on the right side ($12 = 7 + 5$). The third variation also features the empty space on the left side of the equals sign. In this variation we wanted to investigate what numbers students were willing to put in when the number completing the sum disregarding the last number, should be a negative number. We expected this encourage students to view the equation as more of a whole, thereby including the +5 at the end, because negative numbers might be an unacceptable answer or option (Vlassis, 2002).

The context: Matematikfessor.dk an online learning environment for mathematics

In Denmark, as in many other systems, teachers and students increasingly use OLEs. *Matematikfessor.dk*, the environment discussed in this paper, has been running for over 10 years. More than 500,000 students in primary and lower secondary schools have access to the environment and, on a typical day, 45,000 unique students use the variety of tasks offered by the site, and collectively answer 1,500,000 tasks. OLEs like *matematikfessor.dk* therefore have access to a large amount of data and can quickly host replications of tasks such as the ones presented in the sections above in order to generate large amounts of responses. This leads us to the following research question;

What similarities and differences do we see more than 20 years after the original study when implementing the task presented in Falkner et al. (1999) *in an OLE?*

THEORETICAL BACKGROUND

In this section, we collect research about students' conception of the equals sign and comments on the difficulties that emerge from these conceptions. Rittle-Johnson et al. (2011) gives four levels of interpretations of or four meanings to apply to the equals sign in given situations (see table 2).

Level	Description	Core equation structures
Level 4: Comparative relational	Successfully solve and evaluate equations by comparing the expressions on the two sides of the equal sign, including using compensatory strategies and recognizing that performing the same operations on both sides maintains equivalence. Recognize relational definition of equal sign as the best definition.	Operations on both sides with multidigit numbers or multiple instances of a variable
Level 3: Basic relational	Successfully solve, evaluate, and encode equation structures with operations on both sides of the equal sign. Recognize *and generate* a relational definition of the equal sign.	Operations on both sides, e.g.: $a + b = c + d$ $a + b - c = d + e$
Level 2: Flexible operational	Successfully solve, evaluate, and encode atypical equation structures that remain compatible with an operational view of the equal sign.	Operations on right: $c = a + b$ or No operations: $a = a$
Level 1: Rigid operational	Only successful with equations with an operations-equals-answer structure, including solving, evaluating, and encoding equations with this structure. Define the equal sign operationally.	Operations on left: $a + b = c$ (including when blank is before the equal sign)

Table 2. 'Construct Map for Mathematical Equivalence Knowledge' (Rittle-Johnson et al., 2011, p. 3).

One of the central difficulties that students encounter in the transition from an arithmetic thought process to an algebraic one is that they continue to view the equals sign as a 'do something" signal' (Kieran, 1981), or they maintain an urge to 'calculate', out of habit (Alibali et al., 2007). In the context of the task chosen for this study, children do need to be able to consider the right side of an expression involving an equals sign as an expression in its own right. In the words of Rittle-Johnson et al. (2011) an operational view or meaning attached to the equals sign. The main purpose of the task ($8 + 4 = \_ + 7$) is to determine what interpretation of the equals sign a student would apply. In the earlier years in school mathematics students might perceive the equals sign as indication for that calculations has to be made and that the operations on the left side results in a single number on the right side of the equals sign (Alibali et al., 2007; Kieran, 1981).

METHODOLOGICAL CONSIDERATIONS

In August 2020 we implemented the task from Falkner et al. (1999) in the OLE *matematikfessor.dk* as parts of three sets of formative tasks, with a total of 49 unique items about linear equations. The sets were only available for teachers to assign to their students, not for students to find on their own within the environment. A promotion campaign was established in order to notify the teachers subscribing to the services of *matematikfessor.dk* of the formative sets existence and applications. The data (in the form of unique answers) was extracted from *matematikfessor.dk's* database on the 4th of November 2021.

DATA RESULTS

The original task 8+4=__+5

In total 2345 answers were given to the original task presented in Falkner et al. (1999) when we implemented out version in the OLE. In a review of these, we found that only 92 of these answers were from students solving the task multiple times. In table 3 is an overview of the answers the students provided. (64 total answers were omitted. These answers were belonged to a range of 16 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
7	1546	65.9
17	363	15.5
12	343	14.6
3	29	1.2

Table 3: Overview of the answers to the task 8+4=__+5

Answer	Freq	%
7	501	62.9
17	143	17.9
12	119	14.9
3	8	1.0

Table 4: Overview of the answers to the task 8+4=__+5 (age 12 and 13) We examined how 12 to 13 year olds (6th graders) from Denmark answered the task in order to be able to compare with the same age group from the original study. In total 797 students from this age group answered the implementation of the original task. The results can be seen in table 4.

The amount of 12 year olds that gave the answer 7 is 57.3% where the 13 year olds sum up to 64.0%. The average age of the children represented in the data for the original task is 13.97 years, slightly lower than the total average age of 14.08 years of the children represented in all three tasks. See age distribution in figure 1.

The first variation 4+__=7+5

For the second task (the first variation), we received a total of 1203 answers. In a review of these, we found that only 40 of these answers were from students solving the task multiple times. In table 5 is an overview of the answers the students provided. (45 total answers were omitted. These answers were belonged to a range of 14 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
8	996	82.7
3	99	8.2
9	35	2.9
12	27	2.2

Table 5: Overview of the answers to the task $4 + \_ = 7 + 5$

The second variation 6+__=4+5

For the third task (the second variation), we received a total of 824 answers. In a review of these, we found that only 43 of these answers were from students solving the task multiple times. In table 6 is an overview of the answers the students provided. (27 total answers were omitted. These answers were belonged to a range of 13 additional groups of answers that were less than 1% of the answer total)

Answer	Freq	%
3	751	94.9
4	14	1.8
-2	11	1.4
9	11	1.4
2	10	1.3

Table 6: Overview of the answers to the task $6 + \_ = 4 + 5$

Additional results

A total of 351 students have provided answers to all three items. Based on these data the facility of the original task is 69.5%. The facility of the first variation is 83.3% and the facility of the second variation is 93.3%. These students actually represent the overall data very well. Only 32 students have provided two answers to one or more of the items where one of the answers were wrong. We thought it might be interesting to know the exact number of students who either got it wrong first and then right and vice versa. Twenty of the students that provided answers to the task $8 + 4 = \_ + 7$ provided two answers, where the first answer was wrong and the second answer correct. Most of these cases were a situation where either 12 or 17 was the first answer and 7 the second. Five students did in fact provide a correct answer as the first and a wrong answer the second time around.

DISCUSSION

In this section, we discuss the similarities and differences in data results compared to the original studies. In addition, we discuss what possible influence the OLE have on with the similarities and differences. If we compare the data from the original study presented in Falkner et al. (1999) we immediately notice the striking difference in facility among 6^{th} grade students. In the original study, every 6^{th} grade student gave the wrong answer to the task. A later publication (Carpenter et al., 2003) provides additional information about the performance of the task and interpretations made by the authors. The author's comment that the data show that older students are more likely to get the task wrong than younger students are and the author hint at that maybe students get a progressively more operational interpretation of the equals sign based on the teaching at this point in time. Knuth et al. (2006) emphasises that poor performance on measures of understanding the equals sign should not be surprising given the lack of explicit focus in American middle school curricula, although we note that a recent study indicates that American students may have a better understanding of equivalence more generally than some European countries (Simsek et al., 2021). McNeil (2007) finds that performance on equivalence problems such as the ones discussed in this paper decreases with American students from age 7-9 before it increases again from age 9-11. Hence, performance on this item may be particularly influenced by pedagogic and curricular choices. Nonetheless, the data from our study show that students in 6^{th} grade (12-13 year olds) give a correct answer 63% of the time and matches the overall distribution very well. We acknowledge that the original study does not specifically intend to provide information on how 6^{th} grade students perform on a task such as $8 + 4 = \_ + 7$. Rather they intend to provide teachers with a reminder that students' interpretation of the equals sign is of great importance and does not need to be corrected at an older age rather than classroom discussions about the meaning (definition) of the equals sign at lower grades are particular meaningful (Carpenter et al., 2003).

We do get the same wrong answers in our study as in the original. This to some extent prove that the task is not performing in a significantly different way i.e. producing different answers than 20 years ago. We do however wonder why we see the huge difference in the distribution of the answers. 20 years ago in the original study, less than 10% of the participants at every class level gave the answer 7. Now we see a rate of approximately 65%. Granted our data stems from 12-17 years old. With most of the participant being 13-15 (83%). Falkner et al. (1999) mentions that the task was originally carried out by a teacher in a single classroom. When this teacher realized that every student in that classroom provided a wrong answer, she asked her colleagues to use the task with their students resulting in the data in table 1. This means that the observations all stem from the same school. In our study, the data stems from at least 197 schools due to the task being implemented in an OLE. We are however not certain that none of the students in our study received help solving the task. This fact might skew the correct answer percentage towards a higher number. However, it seems unlikely that this should leave us with 60+% correct answers compared to none or almost none. Another obvious difference is nationality of the populations observed in the original study we have American students and in our study the observations stem from Danish students. According to PISA 2018 (https://factsmaps.com/pisa-2018-worldwide-ranking-average-score-of-mathematics-science-reading/) the overall difference in the performance of students in the United States and Denmark is not statistically (or indeed practically) significant. Of course, the task was presented to the American grade 6 students more than 20 years ago and it may be that teachers are now more aware of student's understandings of, and misconceptions about, the equals sign, because of the curricular changes made as a result of the introduction of mathematical competencies in Denmark in 2002. The data collected on the variations of the original task suggests that a similar operational view of the equals sign is being applied even though the empty space is moved to the left side of the equation. This was to be expected, as it is still possible to apply the same operational procedure as the original problem with the empty space on the right side. With the last task, we see an even better performance. The last variation is as expected not similar to the second variation because -2 is not as frequent as the number 3 was in the second variation. This to some extent proves that the choice of numbers matter when designing tasks such as the original task even though the empty space is on the left side of the equals sign. This choice of numbers indicate that students might be more likely to apply a relational interpretation of the equals sign to avoid negative numbers or simply because negative numbers are not accepted in a situation such as this.

CONCLUSION

Based on the differences in the data we believe that, although this task from Falkner et al. (1999), in our opinion is a very good task, the data presented by the authors is not representative of how difficult the task is for 6th grade students. Our data show that the majority of the wrong answers was identical to the ones observed in the original study. This does in our opinion encapsulate one side of the importance of replications studies in mathematics education. On the other hand, our data show a huge deviation from the facility scores of the original study. This is also an important finding for the sake of replication studies in mathematics education. Even though the point of the task presented in Falkner et al. (1999) is not primarily to indicate how difficult it is and present quantitative scores, it is nonetheless important to observe that the scores presented in the original study is an extreme case compared to data collected from a large collection of schools in Denmark 20 years later. With all that said using OLEs to replicate studies such as the performance of the famous task from Falkner et al. (1999) can be great and efficient platforms for achieving additional and in some cases updated information and knowledge.

REFERENCES

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, *18*(1), 37–50.

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A Longitudinal Examination of Middle School Students' Understanding of the Equal Sign and Equivalent Equations. *Mathematical Thinking and Learning*, *9*(3), 221–247.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, *49*(1), 2–8.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically*. Portsmouth, NH: Heinemann.

Falkner, K., Levi, L., & Carpenter, T. (1999). Early Childhood Corner: Children's Understanding of Equality: A Foundation for Algebra. *Teaching Children Mathematics*, *6*(4), 232–236.

Inglis, M., Schukajlow, S., Dooren, W. V., Hannula, M. S., & Leuven, K. (2018). Replication in mathematics education. *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education*, *1*, 2.

Jankvist, U. T., Sánchez Aguilar, M., Misfeldt, M., & Koichu, B. (2021). What to Replicate? *Implementation and Replication Studies in Mathematics Education*, *1*(2), 141–153. https://doi.org/10.1163/26670127-01010015

Kieran, C. (1981). Concepts Associated with the Equality Symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does Understanding the Equal Sign Matter? Evidence from Solving Equations. *Journal for Research in Mathematics Education*, *37*(4), 297–312.

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*(3), 687.

Melhuish, K., & Thanheiser, E. (2018). Reframing replication studies as studies of generalizability: A response to critiques of the nature and necessity of replication. *Journal for Research in Mathematics Education*, *49*(1), 104–110.

Rittle-Johnson, B., Matthews, P., Taylor, R., & McEldoon, K. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, *103*(1), 85–104.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simsek, E., Jones, I., Hunter, J., & Xenidou-Dervou, I. (2021). Mathematical equivalence assessment: Measurement invariance across six countries. *Studies in Educational Evaluation*, *70*, 101046.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. https://doi.org/10.1023/A:1020229023965