Comparing online with paper pencil testing mode when solving stock flow problems using different problem solving strategies

Vivien Lungwitz

Technische Universität Chemnitz
W.-Raabe-Str. 43, 09120 Chemnitz, Germany,
+49-371-531 34143
vivien.roeder@psychologie.tu-chemnitz.de

Keywords: dynamic systems; stock–flow systems; online testing; paper pencil testing; general problem solving; problem solving strategy; gender effect

Introduction: Classic data collection requires participants to come into a laboratory and solve the tasks given on site in a paper pencil format. In past years, it became more and more popular to implement tests via internet. However, it seems not always clear whether results from both methods can be compared that easily. For some fields of research, comparative studies can be found. As of the author's knowledge, no such comparison has been explicitly done for stock flow tasks yet. This study gave five stock flow tasks to two independent samples of participants either in a classic paper pencil setting at university or via internet.
For each setting, about half of the participants were prompted to solve the tasks either with a conventional (CPS) or general (GPS) problem solving strategy. Slight improvements in subjects performance on stock flow tasks were recently observed when participants were primed to use analytical thinking right before being confronted with a stock flow task (Lakeh & Ghaffarzadegan, 2015, 2016) or when they used a GPS strategy instead of a CPS strategy (Röder, 2017). The idea to manipulate the problem solving strategy used by participants before working on stock flow tasks was derived from Youssef-Shalala, Ayres and Schubert, (2014).

This study tried shed light on the following questions: Does it make a difference for the average stock flow performance whether the testing is carried out in a paper pencil setting or via internet? The hypothesis was that it does not make a difference as long as the demographics of the sample are comparable. A second question was, whether prompting the use of a GPS strategy instead of a CPS strategy could boost stock flow performance. Based on previous studies (Röder, 2017; Youssef-shalala et al., 2014) a slight advantage for the GPS was expected. Additionally, the simultaneous use of drawing as well as question based stock flow tasks allowed a direct comparison between different task formats as dependent variables. Finally, the

impact of additional variables that have shown to influence stock flow results in the past, should be calculated (e.g. gender, grade in mathematics, field of study, motivation).

Method: Participants were acquired using mailing lists of different universities in Germany and advertising the studies in social media. The overall process was the same for everybody: After a general information about the study and the agreement to participate voluntarily and anonymous, participants (125 online, 46 paper pencil) read a short introduction on simple dynamic systems and then were given a total of five stock flow problems. Their understanding of how the stock develops was measured by drawing the development of the stock over time (four tasks) and by answering four questions (one task). Both types of tasks were used in past research. Tasks varied in difficulty and had been used in prior studies.

In the general problem solving condition (n = 92), participants were not directly asked for a decision on the development of the stock but rather first prompted to think deeply about the given flow diagram and asked to gather all the information possible before asked about the stock

Results (Fig.1): Overall, the online sample yielded somewhat better results, as did the GPS strategy for the four drawing tasks. Solution rates for the question based task were lower in general and did not produce a clear outcome towards setting or problem solving strategy. Gender had an impact in all groups, favoring men.

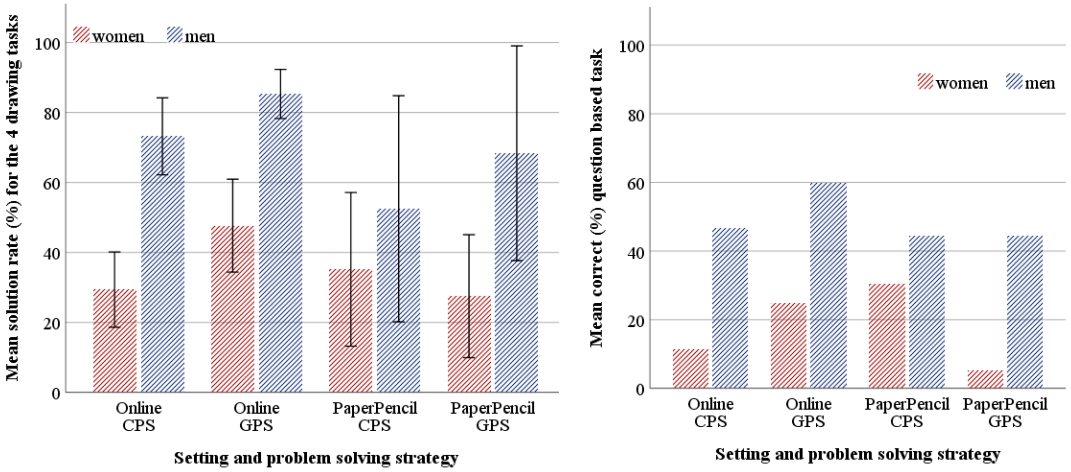

Fig. 1. The average performance in the four drawing tasks (left) and the question based task (right) separated by setting, problem solving strategy and participants' gender. Error bars indicate the 95% confidence interval.

Other measures were collected to see whether the samples drawn differed in regard to the setting and whether those demographic data influenced the results. Multiple regression analysis were calculated for drawing and question based tasks. Gender is the best predictor in both models. For the drawing tasks the manipulation of setting and problem solving strategy also remain highly significant even after controlling for additional variables. Participants' self-rated interest in mathematics also explains a huge deal. The last high school grade in mathematic misses significance on a 5% level but would remain in the model with a more modest alpha cut off. For the question based task the overall predictive value of the regression model is a lot lower than for the drawing tasks. This reflects in only two variables adding significant value to the model: gender and participants' last high school math grade.

Discussion: Concerning the manipulated variables of the study it can be said that this study found an effect for setting (contrary to the prior hypothesis) and a moderate advantage for the general problem solving strategy (in line with the hypothesis). The self-selection bias in the online setting might play an important role in the different results. The rather small sample size in the paper pencil setting makes it difficult to draw far-reaching conclusions.

In addition, the quantitatively different outcomes for drawing and question based tasks raised the question, whether both tasks are comparable and measure the same. Three difficulties in interpreting the results between the different task types are that there was only one question based but four drawing tasks and the question based task was always presented last. Furthermore, the question based task could only be rated as wrong (0%) or correct (100%), whereas the drawing task got a rating between 0 and 100 percent.

Another major finding was a pronounced gender effect on performance in all conditions.

References

Arce-Ferrer, A. J., & Guzmán, E. M. (2009). Studying the Equivalence of Computer-Delivered and Paper-Based Administrations of the Raven Standard Progressive Matrices Test. *Educational and Psychological Measurement*, *69*(5), 855–867. http://doi.org/10.1177/0013164409332219

Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249–286. http://doi.org/10.1002/sdr.198

Brockhaus, F., Arnold, J., Schwarz, M., & Sedlmeier, P. (2013). Does the modification of the representation format affect stock-flow thinking? In R. Eberlein & I. J. Martínez-Moyano (Eds.), *Proceedings of the 31st International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society. Retrieved from www.systemdynamics.org/conferences/2013/proceed/papers/P1253.pdf

Brunstein, A., Gonzalez, C., & Kanter, S. (2010). Effects of domain experience in the stock – flow failure. *System Dynamics Review*, *26*(4), 347–354. http://doi.org/10.1002/sdr

Buchanan, T. (2002). Online Assessment: Desirable or Dangerous? *Professional Psychology: Research and Practice*, *33*(2), 148–154. http://doi.org/10.1037//0735-7028.33.2.148

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, *37*(1), 148–154. http://doi.org/10.3758/BF03206409

Butters, R. B., & Walstad, W. B. (2011). Computer versus paper testing in precollege economics. *Journal of Economic Education*, *42*(4), 366–374. http://doi.org/10.1080/00220485.2011.606087

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, *108*(1), 116–130.

http://doi.org/10.1016/j.obhdp.2008.03.003

Drapeau, C. E., Bastien-Toniazzo, M., Rous, C., & Carlier, M. (2007). Nonequivalence of Computerized and Paper-and-Pencil Versions of Trail Making Test I , ’. *O Perceptual and Motor Skills*, *104*, 785–791. http://doi.org/17688134

Eckford, R. D., & Barnett, D. L. (2016). Comparing paper-and-pencil and Internet Survey methods conducted in a combat-deployed environment. *Military Psychology*, *28*(4), 209–225. http://doi.org/10.1037/mil0000118

Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, *17*(3), 339–346. http://doi.org/10.1016/S0747-5632(01)00002-4

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139. http://doi.org/10.1016/j.chb.2009.10.015

Feldstein, S. N., Keller, F. R., Portman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard versions of the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *13*(3), 303–313. http://doi.org/10.1076/clin.13.3.303.1744

Grieve, R., & Elliott, J. (2013). Cyberfaking: I Can, So I Will? Intentions to Fake in Online Psychological Testing. *Cyberpsychology, Behavior, and Social Networking*, *16*(5), 364–369. http://doi.org/10.1089/cyber.2012.0271

Jensen, E., & Brehmer, B. (2003). Understanding and control of a simple dynamic system. *System Dynamics Review*, *19*(2), 119–137. http://doi.org/10.1002/sdr.267

Jerrim, J. (2016). PISA 2012: how do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy and Practice*, *23*(4), 495–518. http://doi.org/10.1080/0969594X.2016.1147420

Kainz, D., & Ossimitz, G. (2002). Can students learn stock-flow-thinking? An empirical investigation. In P. I. Davidsen, E. Mollona, V. G. Diker, R. S. Langer, & J. I. Rowe (Eds.), *Proceedings of the 20th International Conference of the System Dynamics Society* (pp. 1–34). Albany, NY: System Dynamics Society. Retrieved from http://www.systemdynamics.org/conferences/2002/proceed/papers/Kainz1.pdf

Kapmeier, F. (2004). Findings from four years of bathtub dynamics at higher education institutions in Stuttgart. In M. Kennedy, G. W. Winch, R. S. Langer, J. I. Rowe, & J. M. Yanni (Eds.), *Proceedings of the 22nd International Conference of the System Dynamics Society* (pp. 1–22). Albany, NY: System Dynamics Society.

Kapmeier, F., Happach, R. M., & Tilebein, M. (2014). Bathtub dynamics revisited : Does educational background matter ? In P. Davidsen & E. A. J. A. Rouwette (Eds.), *Proceedings of the 32nd International Conference of the System Dynamics Society* (pp. 1–26). Albany, NY: System Dynamics Society.

Karay, Y., Schauber, S. K., Stosch, C., & Schüttpelz-Brauns, K. (2015). Computer versus paper—Does it make any difference in test performance? *Teaching and Learning in Medicine*, *27*(1), 57–62. http://doi.org/10.1080/10401334.2014.979175

Kasperidus, H. D., Langfelder, H., & Biber, P. (2006). Comparing systems thinking

inventory task performance in German classrooms at high school and university level. In A. Groessler, E. A. J. A. Rouwett, R. S. Langer, J. I. Rowe, & J. M. Yanni (Eds.), *Proceedings of the 24th International Conference of the System Dynamics Society* (pp. 1–28). Albany, NY: System Dynamics Society.

Klovning, A., Sandvik, H., & Hunskaar, S. (2009). Web-based survey attracted age-biased sample with more severe illness than paper-based survey. *Journal of Clinical Epidemiology*, *62*(10), 1068–1074. http://doi.org/10.1016/j.jclinepi.2008.10.015

Le Corff, Y., Gingras, V., & Busque-Carrier, M. (2017). Equivalence of unproctored internet testing and proctored paper-and-pencil testing of the Big Five. *International Journal of Selection and Assessment*, *25*(2), 154–160. http://doi.org/10.1111/ijsa.12168

Korzilius, H. P. L. M., Raaijmakers, S., Rouwette, E., & Vennix, J. A. M. (2014). Thinking aloud while solving a stock-flow task: Surfacing the correlation heuristic and other reasoning patterns. *Systems Research and Behavioral Science*, *31*(2), 268–279. http://doi.org/10.1002/sres.2196

Lakeh, A. B., & Ghaffarzadegan, N. (2015). Does analytical thinking improve understanding of accumulation? *System Dynamics Review*, *31*(1–2), 46–65. http://doi.org/10.1002/sdr.1528

Lakeh, A. B., & Ghaffarzadegan, N. (2016). The dual-process theory and understanding of stocks and flows. *System Dynamics Review*, *32*(3–4), 309–331. http://doi.org/10.1002/sdr.1566

Lungwitz, V., Sedlmeier, P., & Schwarz, M. (2018). Can gender priming eliminate the effects of stereotype threat? The case of simple dynamic systems. *Acta Psychologica*, *188*, 65–73. http://doi.org/10.1016/j.actpsy.2018.05.010.

Mayr, A., Gefeller, O., Prokosch, H. U., Pirkl, A., Fröhlich, A., & De Zwaan, M. (2012). Web-based data collection yielded an additional response bias - But had no direct effect on outcome scales. *Journal of Clinical Epidemiology*, *65*(9), 970–977. http://doi.org/10.1016/j.jclinepi.2012.03.005

Naus, M. J., Philipp, L. M., & Samsi, M. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior*, *25*(1), 1–7. http://doi.org/10.1016/j.chb.2008.05.012

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. Journal of Social Issues, 58, 161–176. http://dx.doi .org/10.1111/1540-4560.00254

Ossimitz, G. (2002). Stock-Flow-Thinking and Reading stock-flow-related Graphs : An Empirical Investigation in Dynamic Thinking Abilities. *The 20th International Conference of The System Dynamics Society*, (May), 1–26. Retrieved from http://www.systemdynamics.org/conferences/2002/proceed/papers/Ossimit1.pdf

Preckel, F., & Thiemann, H. (2003). Online- versus paper-pencil-version of a high potential intelligence test. *Swiss Journal of Psychology*, *62*(2), 131–138. http://doi.org/10.1024//1421-0185.62.2.131

Prisacari, A. A., & Danielson, J. (2017). Rethinking testing mode: Should I offer my next chemistry test on paper or computer? *Computers and Education*, *106*, 1–12. http://doi.org/10.1016/j.compedu.2016.11.008

Phuah, T. L. K. (2010). Can people learn behaviours of stock and flow using their ability to calculate running total ? An experimental study. In *28th International Conference of the System Dynamics Society* (pp. 1–19). Seoul, Korea. Retrieved from https://bora.uib.no/handle/1956/4171?show=full

Qi, L., & Gonzalez, C. (2015). Mathematical knowledge is related to understanding stocks and flows: Results from two nations. *System Dynamics Review*, *31*(3), 97–114. http://doi.org/10.1002/sdr.1539

Röder, V. (2017). The Attempt to Boost Stock Flow Performance by Using a General Problem Solving Strategy and a Reappearing Gender Effect. In *Proceedings of the 35th International System Dynamics Conference*. Cambridge, MA.

Saleh, M. (2000). The hard core of the system dynamics research programme. In P. I. Davidsen, D. N. Ford, & A. N. Mashayekhi (Eds.), *Proceedings of the 18th International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society. Retrieved from www.systemdynamics.org/conferences/2000/PDFs/saleh226.pdf

Schwarz, M. (2016). *Vom stockenden Verständnis fließender Zusammenhänge: Darstellungs- und personenbezogene Einflussfaktoren auf das basale Verständnis einfacher dynamischer Systeme*. TU Chemnitz. Retrieved from http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-203692

Schwarz, M., Epperlein, S., Brockhaus, F., & Sedlmeier, P. (2013). Effects of illustrations, specific contexts, and instructions: Further attempts to improve stock–flow task performance. In R. Eberlein & I. J. Martínez-Moyano (Eds.), *Proceedings of the 31st International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society. Retrieved from http://www.systemdynamics.org/conferences/2013/proceed/papers/P1168.pdf

Smith, B., Smith, T. C., Gray, G. C., & Ryan, M. A. K. (2007). When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. *American Journal of Epidemiology*, *166*, 1345–1354. http://dx.doi.org/10.1093/aje/kwm212

Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501–531. http://doi.org/10.1002/sdr.261

Sterman, J. D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review*, *26*(4), 316–334. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/sdr.447/abstract;jsessionid=116FB524DF2A9D10857F2ABEB799BF53.f01t02

Vanpatten, B., Trego, D., & Hopkins, W. P. (2015). In-Class vs. Online Testing in University-Level Language Courses: A Research Report. *Foreign Language Annals*, *48*(4), 659–668. http://doi.org/10.1111/flan.12160

Veldhuis, G., & Korzilius, H. P. L. M. (2012). Seeing with the mind - The role of spatial ability in inferring dynamic behaviour from graphs and stock and flow

diagrams. In E. Husemann & D. Lane (Eds.), *Proceedings of the 30th International Conference of the System Dynamics Society* (pp. 1–23). Albany, NY: System Dynamics Society.

Youssef-shalala, A., Ayres, P., & Schubert, C. (2014). Using a General Problem-Solving Strategy to Promote Transfer. *Journal of Experimental Psychology: Applied*, *20*(3), 215–231. http://doi.org/10.1037/xap0000021