Electronic Supplementary Material (ESI) for Chemistry Education Research and Practice.

Supplementary Material

I Learning Activities

German versions of drawing-assisted learning activities according to SIMMS can be found on the following website: https://simms-uni-bielefeld.de

II Evaluation of affective Factors

A Methods to evaluate affective factors

Questionnaire on intrinsic motivation

Beside cognitive factors, motivation has a major influence on student learning. In order to detect possible motivational differences between the groups and test the comparability of the two simulation-based variants, a short scale on intrinsic motivation has been used (KIM, Krombass and Harms, 2006; Wilde *et al.*, 2009). This instrument, based on self-determination theory by Deci and Ryan (1975; Devetak, 2011) encompasses twelve items to assess the four constructs interest/pleasure, perceived competency, perceived autonomy as well as pressure/strain on a Likert scale featuring five steps. The interest/pleasure subscale is considered a self-report measure of intrinsic motivation. The second and third factors are considered predictors of intrinsic motivation, the factor strain/pressure constitutes a negative predictor for situational intrinsic motivation (Wilde *et al.*, 2009). Reliability of the KIM scale in this study yielded Cronbach's $\alpha = .76$ for *interest/pleasure*, Cronbach's $\alpha = .86$ for *perceived competence*, Cronbach's $\alpha = .89$ for *perceived autonomy* as well as Cronbach's $\alpha = .70$ for *pressure*. Accordingly, the questionnaire can be classified as reliable.

Quantitative analysis of KIM short scale:

Students' intrinsic motivation was assessed after the implementation of the interventions using the Short Scale of Intrinsic Motivation (KIM, Wilde *et al.*, 2009) on a 5-point Likert scale. To determine statistical differences between the groups with regard to intrinsic motivation, the respective scores of the items on the four constructs of the short scale were first averaged (scoremax = 5). The resulting interval-scaled variables were then subjected to multivariate analysis of variance (MANOVA). MANOVA was used because the subscales are dependent variables that correlate with each other. Group differences are examined here simultaneously across several dependent variables. The use is reasonable to control the alpha error (Eschweiler *et al.*, 2007). In all cases, the analysis was undirected and the significance level was set at $\alpha = .05$.

B Results on affective factors

Since comparing performances between single-choice-tasks and drawing tasks might be distorted by affective factors during intervention, students' intrinsic motivation has been assessed via the KIM short scale (Krombass and Harms, 2006; Wilde *et al.*, 2009), addressing the constructs *interest*, *perceived competency*, *perceived freedom of choice* and *pressure*. Constructs are rated with score values between 1 (scoremin) and 5 (scoremax). Average values and standard deviations for all groups can be found in table 9 and figure 14.

Table 9: Mean values and standard deviations of affective constructs measured by the KIM short scale for all three groups

construct	drawing group M (SD)	single-choice group M (SD)	control group M (SD)
interest	3.38 (0.63)	3.24 (0.80)	3.29 (0.69)
p. competency	2.92 (0.70)	2.87 (0.91)	3.10 (0.93)
p. autonomy	3.50 (1.04)	3.46 (1.14)	2.70 (0.72)
pressure	2.43 (0.88)	2.35 (0.79)	2.02 (0.89)

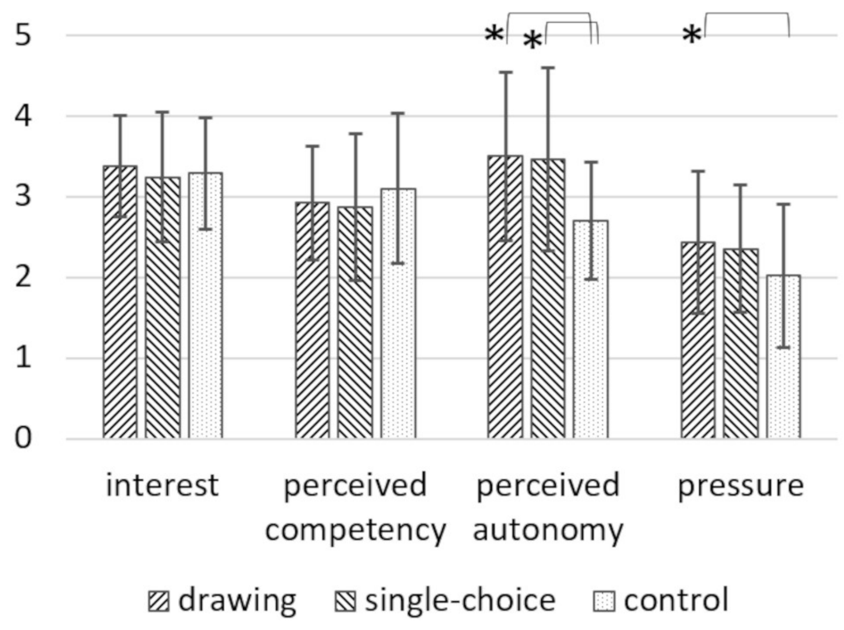

Fig. 14. Comparison between all three groups concerning affective factors assessed by the KIM short scale.

A one-way MANOVA revealed a statistically significant difference between the groups on the combined dependent variables ($F(8,322) = 3.898$, $p = <.001$, $\eta^2 = .088$). Homogeneity of variances was asserted

using Levene's Test, which showed that equal variances could be assumed for interest (p = .423), perceived competency (p = .095) and pressure (p = .408), but not for perceived autonomy (p = .002).
Post-hoc univariate ANOVAs were conducted for every dependent variable. Results show a statistically significant difference between the groups for perceived autonomy ($F(2,163) = 10.095$, $p < .001$, partial $\eta^2 = .110$) and pressure ($F(2,163) = 3.243$, $p = .042$, partial $\eta^2 = .038$), but not for interest ($F(2,163) = 0.599$, $p = .550$, partial $\eta^2 = .007$) and perceived competency ($F(2,163) = 1.005$, $p = .368$, partial $\eta^2 = .012$).
Tukey HSD post-hoc analysis for the subscale *pressure* revealed a significant (p = .038) difference between the drawing group and control group (M_{Diff} = 0.41, 95%-CI[0.02, 0.79]).
Games-Howell post-hoc analysis for the subscale *perceived autonomy* revealed a significant (p <.001) difference between the drawing group and control group (M_{Diff} = 0.80, 95%-CI[0.41, 1.19]) and furthermore a significant (p <.001) difference between the single-choice group and control group (M_{Diff} = 0.76, 95%-CI[0.30, 1.22]).
Differences can be found, but these are only between a treatment and a control group and not between the treatment groups. Thus, a comparability of the two treatment groups is given.

C Discussion on affective factors

It is not the focus of this work to investigate the influence of affective factors in simulation-based learning. However, a comparison of affective factors might reveal differences between the groups and/or interventions unforeseen by the researchers.
For both treatment groups all four affective constructs assessed by the KIM short scale are quite similar. Hence, differences in learning progression between the two treatment groups are not distorted by variances in affective factors. Comparing both treatment groups to the control group, two significant differences have been detected, revealing a higher perceived autonomy for both treatment groups as well as lower pressure for the control group. While not suggesting some sort of causality between these affective differences and students' post-performances, we discuss them from an instructional point of view.
We don't find the relatively high pressure in both treatment groups surprising and assume them to stem from a combination of individual settings and cognitively challenging tasks, putting the pressure on one's own ideas rather than collaborating with classmates and joining ideas as well as forces to work on constructing mental models. This might be alleviated once these activities are worked on collaboratively. As for autonomy, we are surprised to find significantly higher values for both treatment groups, since the simulation-based instructional concept SIMMS is quite rigid and features strict guidance. However, while the linear learning path itself provides little leeway, allowing students to interactively explore simulations on their own and in their own time seems to be enough to surpass perceived autonomy in regular lessons.

D Literature

Deci E. L. and Ryan R. M., (1975), *Intrinsic Motivation and Self-Determination in Human Behavior*, Springer.

Devetak I., (2011), Teachers' influence on students' motivation for learning science with understanding. *Prog. Educ.*, **19**, 77–103.

Eschweiler M., Evanschitzky H., and Woisetschläger D., (2007), Ein Leitfaden zur Anwendung varianzanalytisch ausgerichteter Laborexperimente [A guide to the application of variance-analytically oriented laboratory experiments], in *WiSt - Wirtschaftswissenschaftliches Studium [Economic studies]*, Berthold N. and Lingenfelder M. (eds.)., Beck&Vahlen, pp. 546–554, DOI: 10.15358/0340-1650-2007-12.

Krombass A. and Harms U., (2006), [A computer-based information system on biodiversity as a motivating and learning supplement to the exhibits of a natural history museum]. *Zeitschrift für Didakt. der Naturwissenschaften*, **12**, 7–22.

Wilde M., Bätz K., Kovaleva A., and Urhahne D., (2009), Überprüfung einer Kurzskala intrinsischer Motivation (KIM) [Testing a short scale of intrinsic motivation (KIM)]. *Zeitschrift für Didakt. der Naturwissenschaften*, **15**, 31–45.