INTEGER INSTRUCTION: AN EXPERIMENTAL COMPARISON

Andreas Koukkoufis and Julian Williams

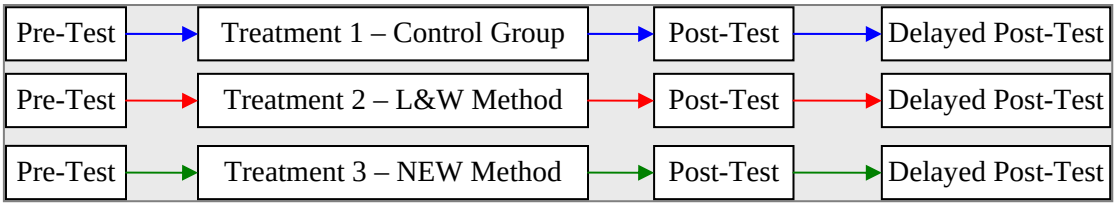
The University of Manchester, School of Education

Two versions of Linchevski & Williams' 'dice games' method for integer addition and subtraction were experimentally contrasted. We describe the methods, present some statistical analyses and discuss the findings. We finally suggest more attention to situated meanings and intuitions in realistic contexts is needed.

INTRODUCTION

Linchevski & Williams (1999) in the *dice games* method for integer addition and subtraction targeted students' intuitive reification of integers, capitalizing on situated intuitions – a key to the intuitive instruction of integer addition and subtraction. As part of an ongoing PhD, a quasi-experimental design is examined where the original *dice games* method (Linchevski & Williams, 1999) is *L&W method* and a variation of it is *NEW method*. The two methods were contrasted with a control group (no integer addition and subtraction instruction), as shown in figure 1:

Figure 1: The quasi-experimental design



The students completed a pre-test, a post-test and a delayed post-test which (via Rasch analysis) produced three repeated measures for each student on an integer ability (IA) scale. We present the differences between L&W method and NEW method followed by some results of their statistical analyses based on the IA measures and discuss the consequences.

THE TWO EXPERIMENTAL METHODS (L&W AND NEW METHOD)

As L&W method (the original *dice games*) is detailed in Linchevski & Williams (1999) and in Koukkoufis & Williams (in press), here we stress the differences. Both contain 4 games that students play in groups of 4, divided in two teams: yellow and red team for L&W method; teams 1 and 2 for NEW method. In both methods, students throw in turns: in game 1, a yellow and a red die; in game 2, these and an *add/sub die* (i.e. a die giving *add* or *sub* (subtract)); in game 3, only an *integer die* (scoring -3, -2, -1, +1, +2, +3) replacing the red and yellow die; in game 4, the *integer die* and the *add/sub die*. Each group has two abacuses where their points are recorded. In L&W method, each abacus is used by a yellow team and a red team student recording points for both teams. In NEW method, each abacus is used only by one team for recording points for that team. Yet, the dice have different meanings in the methods.

In L&W method, the yellow and red die results are *points for the yellow* and *for the*

red team respectively. The team that gets 8 points ahead of the opponent wins. Despite the teams only *getting* points (not losing any), as they try to get ahead, a point for the yellows is a point less for the reds. The students must understand that the dice scores can be cancelled (*cancellation strategy*) and that adding a point to one team is equal to taking a point from the other (*compensation strategy*) as these strategies are fair in the games. This *intuition of fairness* is crucial. Equally vital is in game 3 the transition from points *for the yellow* to *pluses* (i.e. *3 for the yellow* is *+3*) and from points *taken from* the yellow (or *points for* the red) to minuses (i.e. *3 for the red* is *-3*). Similarly, *+3* is *3 taken from the red*: positives and negatives have opposite meanings for the teams.

In NEW method, yellow and red points are *winning* and *losing points* respectively for both teams. Thus, the teams are *both winning and losing* points. To win, a team needs an overall score of 8 winning points (e.g. if a team has 2 losing and 10 winning points, its overall score is 8 winning points). Based on the game *intuition of fairness* (as in L&W method), but now also based on *winning and losing intuitions*, the students here too construct a *cancellation* and a *compensation strategy*. In game 3 winning points become *pluses* and losing points become *minuses* for *both* teams (e.g. +3 is 3 winning points): here integers have the same meanings for both teams.

We hypothesize that the different meanings of the yellow and red die and the additional winning and losing intuition in NEW method are at the heart of the methods' differences. First, negative integers in L&W method are not actually sub-zero quantities: they are the scores of the red team and therefore positives and negatives have the *same qualities* (they are above zero), but are *opposite amounts* because of the game context. We believe this can cause some confusion with integer comparison and reduce the effectiveness of L&W method. In contrast, in NEW method positives and negatives have opposite qualities, as negatives are below zero (they are *losing points*). We believe NEW method students will not experience the same problem. Secondly, in NEW method the students have the extra intuition of winning and losing to build new knowledge upon. As it is not just an intuition limited to the games context (winning and losing are opposite in almost all aspects of life), we hypothesize that it will allow more students to make integer operations intuitive. Further, we assume that if NEW method students are confused at some point (before or after instruction completion), it will be easier for them to reconstruct knowledge based on this game intuition than just on the intuition of fairness. Finally, because of the double meanings of integers in games 3 and 4 in L&W method, Linchevski and Williams (1999) report a counter-intuitive situation when negative integers *add* to the red team. We believe this counter intuitive moment will not exist in NEW method.

Due to these differences in the situated meanings of integers and in the situated intuitions, we hypothesize NEW method students will improve their integer abilities more, while L&W method students will improve their abilities also but to a lesser extent. In this paper we ask the research question:

RQ: What are the effects of the 2 experimental methods in students'

attainment of integer conceptions?

ANALYSIS OF EXPERIMENTAL DATA THROUGH SAMPLE MEANS

In our experimental design year 5 students from one Greater Manchester school were examined, a class for each treatment. A summary of students' integer abilities by treatment and test type is presented in table 1.

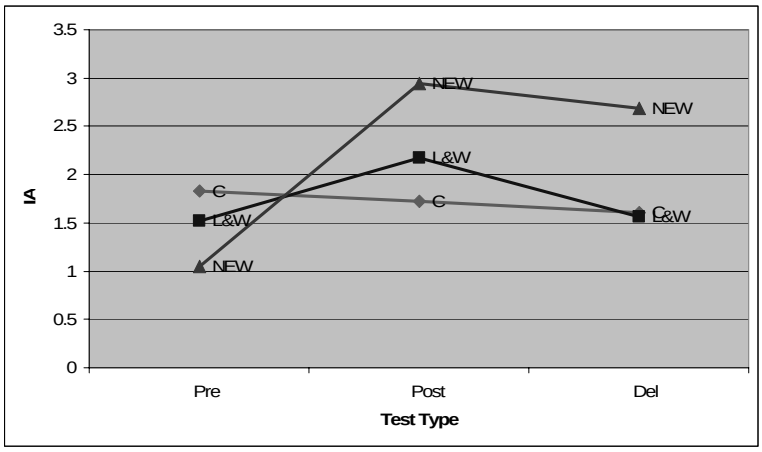
Table 1: Summary of students' IA measures by treatment and test type

	L&W Method			NEW Method			Control Group		
	Pre	**Post**	**Del**	**Pre**	**Post**	**Del**	**Pre**	**Post**	**Del**
Valid Measures	15	15	15	21	21	21	23	23	23
Total Cases	15	15	15	21	21	21	23	23	23

In the table we notice that the sample sizes for the three treatments by test type are quite small, reflecting limited access and the time-consuming design of the experiment (small group teachings; 4 one-hour meetings per group for teaching, testing and interviewing). This is a limitation for our experiment because some differences may not be found to be statistically significant. Yet, if they are found, we should not worry about the small sample sizes for these differences: there is a danger that we may not find a statistically significant difference, but there is not a danger that a statistical significance does not exist. Another limitation is that the small-group instructions allow increased attention to each child and favour intense learning. This limitation affects the comparison of L&W method and NEW method with the control group, but not the comparison between methods L&W and NEW.

Using the IA measures, the experimental samples were examined using descriptive statistics. Here we will only examine the sample means, represented in figure 2.

Figure 2: The sample means by treatment and test type



From figure 2, we see that in the pre-test, NEW method has the lowest mean, L&W method is 0.47 logits higher and the control group has the highest mean (higher of NEW method by 0.79 logits and of L&W method by 0.32 logits).

In the post-test, the control group mean remained almost unchanged (decrease of 0.11 logits – effect size (ES) = -0.10). In contrast, L&W method presented an increase of

0.66 logits (ES = 0.35). The change in NEW method was more impressive: it presented an increase of 1.9 logits (ES = 1.1). This great increase not only covered the pre-test difference between NEW method and the control group, but NEW method had a higher mean than the control group by 1.2 logits. Finally, though in the pre-test NEW method had a lower mean than L&W method, in the post-test NEW method mean was higher than L&W method mean by 0.77 logits. Thus, for the samples, NEW method outperformed both L&W method and the control group between the pre-test and the post-test, while L&W method also did better than the control group.

In the delayed post-test, again the mean of the control group remained almost unchanged (decrease of 0.12 logits – ES = -0.15) in comparison to the post-test, but this did not apply for the two experimental groups. L&W method presented a decrease of 0.61 logits (ES = -0.32). Further, though the delayed post-test mean of L&W method remained 0.05 logits higher than the corresponding pre-test mean (ES = 0.03), this difference is so small that one can hardly claim that eventually the acquired knowledge was retained. On the other hand, in NEW method the sample mean presented a smaller decrease of 0.26 logits (ES = -0.16). Thus, for NEW method the delayed post-test mean is higher than the corresponding pre-test score by 1.6 logits (ES = 1.3). As a result of this decrease, the gap between the NEW method and the control group became 1.1 logits. Finally, NEW method mean is higher than L&W method mean by 1.1 logits.

INVESTIGATION OF GENERALIZATIONS TO THE POPULATION

To examine the generalisations to the population regarding the three treatments, we have constructed (through the use of the software 'R') the following model:

Integer Ability (IA) = Treatment + TestType + Treatment X TestType + Subject

In this model: (a) Integer Ability (continuous variable) is the ability measured by the scale we have constructed through Rasch analysis; (b) Treatment (unordered categorical variable with 3 levels) refers to the control group, L&W method and NEW method; (c) TestType (ordered categorical variable with 3 levels) refers to the pre-test, the post-test and the delayed post-test; (d) Treatment X TestType is the interaction of the two above factors, which allows the comparison between the factors; and (e) Subject (unordered categorical variable) indicating the students participating in the experiment. The inclusion of the students in the model allows us to construct a more appropriate model for our repeated measures design. Regarding the use of regression models for repeated measures designs, see Hutcheson & Sofroniou (1999).

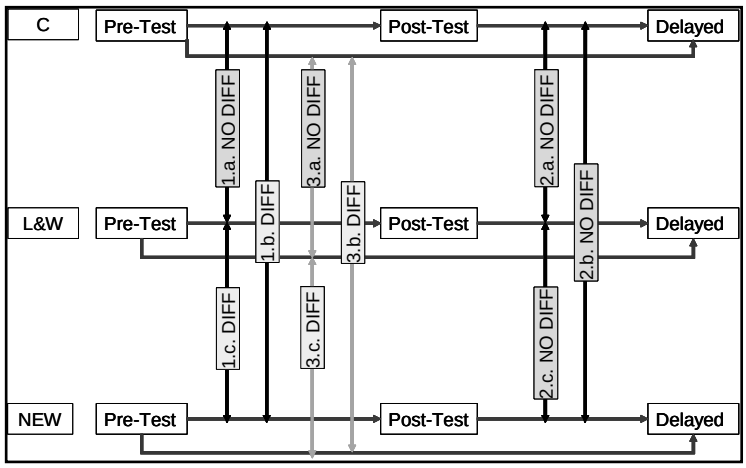
The produced model presented $R^2 = 0.66$, meaning that the model predicts about 66% of the variance of the dependent variable. As the F-statistic gave a p-value of 6.494e-09, we can say confidently that the linear relationship is significant. Based on these two characteristics of the model, we conclude that the model is appropriate and therefore can be used for statistical inferences about the population.

Based on the model, the null hypotheses that there were no statistically significant differences between the following differences were examined:

1. The difference of the means of integer ability (IA) in the pre-test and the post-test for: a. the control group and L&W method (CPost - CPre = L&WPost - L&WPre), b. the control group and NEW method (CPost - CPre = NEWPost - NEWPre), and c. L&W method and NEW method (L&WPost - L&WPre = NEWPost - NEWPre).
2. The difference of the means of IA in the post-test and the delayed post-test for: a. the control group and L&W method (CDel - CPost = L&WDel - L&WPost), b. the control group and NEW method (CDel - CPost = NEWDel - NEWPost), and c. L&W method and NEW method (L&WDel - L&WPost = NEWDel - NEWPost).
3. The difference of the means of IA in the pre-test and the delayed post-test for: a. the control group and L&W method (CDel - CPre = L&WDel - L&WPre), b. the control group and NEW method (CDel - CPre = NEWDel - NEWPre), and c. L&W method and NEW method (L&WDel - L&WPre = NEWDel - NEWPre).

The answers to these hypotheses were schematized with arrows in figure 3:

Figure 3: The comparisons for the hypotheses schematised



As ‘DIFF’ means that there is a statistically significant difference and ‘NO DIFF’ means that a not statistically significant difference was found:

1. In all the comparisons between the control group and L&W method no significant differences were found (hypotheses 1.a., 2.a. and 3.a.). However, this may be a small-sample effect. This is a limitation of our small sample sizes.
2. In the comparisons between the control group and NEW method, the differences in hypotheses 1.b. and 3.b. were statistically significant, but for hypothesis 2.b. it was not.
3. In the comparisons between L&W method and NEW method, the differences in hypotheses 1.c. and 3.c. were statistically significant. For hypothesis 2.c. the difference was not statistically significant. This may be a small-sample effect, but we cannot tell. Therefore, this is a limitation to our analysis.

DISCUSSION

We have presented the differences of two versions of the *dice games*, which we called L&W method and NEW method. Our analyses have shown that both methods improve students' integer abilities between the pre-test and the post-test, though the improvement for L&W method was not statistically significant – we suggest due to the small samples. Further, for both methods students' abilities decrease between post-test and delayed post-test, but for L&W method the decrease is so large that almost all the new knowledge seems to have been lost, whereas in NEW method most of the new knowledge is retained. In both cases the decreases were not statistically significant, but we suspect that for L&W method this is again a small-sample effect. Based on all these evidence, we can be confident that NEW method was more effective to L&W method and to the control group. Regarding L&W method, we cannot say for certain that it is preferable to the control group but we assume it should be, because in this method too the students produce intuitive strategies (Koukkoufis & Williams, 2006) and they have shown an increase of 0.66 logits (ES = 0.35) between the pre-test and the post-test which is noteworthy. To be sure, a comparison of L&W method with the control group with larger numbers of students would be needed.

Concluding, we believe that this experiment – in addition to the better learning outcomes produced and retained through NEW method – enforces the significance of situated meanings (e.g. the situated meanings of integers in our case) and situated intuitions (e.g. here, the intuitions of fairness and of winning and losing) in realistic contexts. Particularly, we suggest that the changes in these game characteristics in NEW method have led to a chain of differences in the games and explain these significant differences in their learning outcomes. We take advantage of this opportunity to call on the need for more attention on situated meanings and intuitions in realistic contexts.

REFERENCES

Hutcheson, G., & Sofroniou, N. (1999). The Multivariate Social Scientist. London: Sage Publications.

Koukkoufis, A., & Williams, J. (2006). Integer Instruction: A Semiotic Analysis of the Compensation Strategy. Paper presented at the PME30, Prague.

Koukkoufis, A., & Williams, J. (in press). Semiotic Objectifications of the Compensation Strategy: En Route to the Reification of Integers. Revista Latinoamericana de Matemática Educativa.

Linchevski, L., & Williams, J. (1999). Using intuition from everyday life in 'filling' the gap in children's extension of their number concept to include the negative numbers. Educational Studies in Mathematics, 39(1-3), 131-147.