

# When was randomisation first used in educational research? A brief historical methodological perspective

**Carole Torgerson**

University of York, UK

**Corresponding author:** Carole Torgerson. Email: [carole.torgerson@york.ac.uk](mailto:carole.torgerson@york.ac.uk)

## Introduction

This brief note aims to add to the debate in the literature around the first use of randomisation in educational research, specifically in relation to previously published articles by Forsetlund et al.<sup>1</sup> and Hedges and Schauer<sup>2</sup> in which they claim the first use of randomisation was in 1931 or 1919. This historical, methodological perspective analyses in detail the 1919 contender for the first use of randomisation by Hedges and Schauer<sup>2</sup> (published in Styles and Torgerson<sup>3</sup>). This is the monograph by Cummings: *Improvement and the Distribution of Practice*.<sup>4</sup>

In summary, early milestones in the history of randomised trials in education include *possibly* the first pragmatic experiment, by Pearson<sup>5</sup>; *possibly* the first randomised experiment, by Cummings<sup>4</sup>; *probably* the first use of alternation, by Remmers<sup>6</sup>; and *probably* the first two randomised experiments, by Walters.<sup>7,8</sup> Each of these is described below.

## Pearson

In 1911, Henry C. Pearson undertook possibly the first pragmatic experiment in education.<sup>5</sup> Torgerson and Torgerson uncovered this study, which used a ‘comparative’ type of experimentation, while researching for their article on pragmatic experimentation in educational research.<sup>9</sup> It was a controlled experiment in the teaching and learning of spelling of homonyms and recognised the need for a comparative type of experimentation to reflect real-world practice. The chief defect of the laboratory or explanatory experiment, according to Pearson, was that it ‘isolates from its natural setting the issue to be tested’ and should instead be ‘surrounded by the normal accompaniments of its classroom situation’.<sup>5</sup> His experiment aimed to determine which of two methods was more effective in teaching homonyms, by pre-testing children in two classrooms from each of five grade levels (grades III to VII; ten classrooms in all) and then assigning one class of each grade to one of the two methods (teaching homonyms

side by side or with an interval of 3 to 4 days), followed by post-testing of all children. The homonyms selected for the interventions were increasingly difficult as the age of the children increased. This was a cluster trial,<sup>10</sup> although it was analysed at the level of the pupil, except in the case of the data for Grade V, where the mean, average and average deviation for each class were presented. Although this analysis only presented these values for two classes (two clusters with one class in each cluster, i.e., a two-class experiment), it gave a foretaste of the appropriate analysis for cluster randomised trials by using cluster-level data. This method was described in more detail by Lindquist<sup>11</sup> in his textbook discussing an appropriate analytical technique for data generated by the use of cluster randomisation, which is now largely the norm for educational experiments. It also appears that Pearson used change scores to make the comparison between the two classes in each grade in terms of (a) average decrease of errors between pre- and post-test, (b) average improvement sub-divided by the number of initial errors and (c) sub-divided by ‘good’ and ‘bad’ spellers. There is no evidence in this article that randomisation was used to allocate the two intervention classes to conditions within each grade. However, despite this limitation, strengths include the fact that Pearson acknowledged the small sample size of the experiment (although he stated that if the experiment were to be replicated with a larger sample, he did not think this would change the results); teacher effect was acknowledged; and a replication trial was undertaken and reported in the same article.

## Remmers and Walters

In 2007, Forsetlund and colleagues reported their systematic search of the literature between 1867 and 1948 for the first reported use of randomisation in social and educational studies.<sup>1</sup> They identified nine potentially eligible studies, including the earliest study by Remmers,<sup>6</sup> and subsequently argued for Jack Walters’ counselling pilot experiment of 1931 and his replication trial of the following year as strong contenders for the first randomised

trials in education. This was due to the experiment by Remmers,<sup>6</sup> having clearly used alternation and not randomisation as the method of allocation. Walters, on the other hand, clearly stated that he used random allocation for the pilot and replication experiments (see below).

## Cummings

Robert Alexander Cummings' monograph *Improvement and the Distribution of Practice* was published in 1919.<sup>4</sup> It reports a series of pragmatic experiments undertaken at the teachers' college of Columbia University, USA, comparing practice periods 'Equal' in length and practice periods 'Reducing' in length with children and adults in authentic settings in various subjects, for example, mathematics, history. One of these experiments – the Ohio experiment – was uncovered by Larry Hedges and Jake Schauer in their article 'Randomised trials in the USA'.<sup>2</sup> This was one of six articles on methodological debates, questions and challenges about randomised trials that were published in the 2018 Special Issue of the *Education Research* journal.<sup>3</sup> Hedges and Schauer give a succinct overview of the history of educational experiments, from the first recorded randomised trial through five historical periods to the state of the art in the 21st century.

The first of Hedges and Schauer's five historical periods covered the period of Pearson's, Cummings' and Walters' early experiments – all of which were pragmatic – although most controlled trials in this period were laboratory experiments. In their article, Hedges and Schauer introduced several earlier studies undertaken by Cummings as contenders for the first randomised trial in education (in 1919), which was more than 10 years' earlier than the first studies published by Walters in the early 1930s.

In Cummings' 1919 monograph, there are three pieces of evidence about the method of allocation to intervention and control groups. The first mention of random selection is on page 51, as cited in Hedges and Schauer's article: 'The factor of the teacher was equalised by a random selection of the classes which made up the two groups'. This *could* mean that the classes were randomised, *or* the teachers were randomised to the classes, *or* the classes in the sample were a *random sample* of classes from the schools. Cummings then discusses the issue of lack of baseline balance in the factor of previous training and states, 'the random method of selecting the pupils would tend to favour one group as much as the other'. This seems to suggest that the pupils were randomised (which seems unlikely), but it could mean the pupils were a random sample of the pupils in the school because the classes were a random sample of the classes in the school (see previous quotation). However, on the previous page is evidence not cited in Hedges and Schauer's article but, by inference, is the information used to lead to their conclusion that randomisation was at the level of school:

The *Equal* and *Reducing* groups were made up from the pupils of the seven villages as follows: The *Equal* group included all the classes at Rocky Ridge, Lakeside, and Greenwich, and grades 3, 5 and 7 from Oak Harbor. The *Reducing* group included grades 6 and 8 from Oak Harbor and all the classes at Elmore, Waterville, and Weston.

This *could* mean that randomisation was at the level of school (although the word random is not used here) and only six of the seven schools could have been randomised at this level – the other school was randomised by class. Thus, we cannot be sure that this is a randomised trial.

In another of Cummings' experiments, he uses the phrase pupils 'taken at random as ours were' to justify the assumption that one factor (previous practice) at baseline was equalised. However, it is not clear whether this is randomisation of pupils or of classes. Also, if this was randomisation of *classes*, the fact that all but one of the classes are in alphabetical order suggests that this was *not* randomisation.

As an aside, in these early controlled experiments, Cummings appears not to have used correct methods of analysis. On page 12, we learn that in the Lyndhurst experiments, intention-to-treat (ITT) analysis (the correct analysis)<sup>12,13</sup> was not used. In his experiment, Cummings took two groups *unequal* on one factor and used a method of removing outliers to *equalise* the groups on this factor (initial ability), and he called this the 'pairing off method'.

## Summary and conclusions

It is likely that randomisation was used in educational research more widely than suggested by the references cited here. However, it is difficult to identify published research that unambiguously states randomisation was used to produce equivalent groups. In Linn's book, some of the language used implies that randomisation was widely used; however, there is little evidence to support that contention. Early trials tended to use the phrase 'random sampling' to mean randomisation, which may mean some studies have been overlooked. For instance, Walters<sup>7</sup> states 'Freshman were divided into two groups by random sampling' rather than using current terminology where we might say 'Freshman were randomised into two groups'. It is likely that Walters *did* use randomisation to form the two groups as the population of interest ('delinquent' Freshman) was specified and the whole population was allocated using random sampling, thereby producing two equivalent groups. The Cummings' terminology, while it *could have* been referring to randomisation, left sufficient doubt in the description to mean it is likely to be a quasi-randomised trial. Nevertheless, the literature cited in this paper shows that the design of modern trials owes a great deal to these pioneers of controlled experiments. The use of alternation, while somewhat frowned upon today, is a legitimate technique to form

comparable groups *if* properly used; however, experience shows that its implementation is often sub-optimal, which has the potential to introduce selection bias. Pearson's desire to undertake an experiment in classrooms rather than in the confines of a laboratory paved the way for large pragmatic experiments, or field trials, replicating the real-life educational experience of students and teachers to become the norm of educational experiments today. To conclude, many educational researchers were exploring techniques to reduce the role of bias on the interpretation of the results of experimental research in the first half of the 20th century.

## Declarations

**Competing Interests:** None declared.

**Funding:** None declared.

**Ethics approval:** Not applicable.

**Guarantor:** CT.

**Contributorship:** Sole author.

**Provenance:** Invited article from the James Lind Library.

## References

1. Forsetlund L, Chalmers I and Bjorndal A. When was random allocation first used to generate comparison groups in experiments to assess the effects of social interventions? *Econ Innov New Technol* 2007; 16: 371–384.
2. Hedges L and Schauer J. Randomised trials in education in the USA. In Styles B and Torgerson C (eds.) Randomised controlled trials (RCTs) in education research – methodological debates, questions, challenges. *Educ Res* 2018; 60: 265–275.
3. Styles B and Torgerson C. Randomised controlled trials (RCTs) in education research – methodological debates, questions, challenges. *Educ Res* 2018; 60: 255–264.
4. Cummings RA. *Improvement and the Distribution of Practice*. New York: Teachers College, Columbia University, 1919. See <https://www.loc.gov/resource/gdcmassbookdig.improvementdistr01cumm/?st=gallery> (last checked 5 August 2025).
5. Pearson HC. The scientific study of the teaching of spelling. *J Educ Psychol* 1911; 2: 241–252.
6. Remmers HH. A diagnostic and remedial study of potentially and actually failing students at Purdue University. *Bull Purdue Univ Stud Higher Educ* IX 1928; 28: 1–164.
7. Walters JE. Seniors as counselors. *J Higher Educ* 1931; 2: 446–448.
8. Walters JE. Measuring effectiveness of personnel counseling. *Personnel J* 1932; 11: 227–236.
9. Torgerson CJ and Torgerson DJ. The need for pragmatic experimentation in educational research. *Econ Innov N Technol* 2007; 16: 323–330.
10. Moberg J and Kramer M. A brief history of the cluster randomised trial design. *J R Soc of Med* 2015; 108: 192–198. See <https://www.jameslindlibrary.org/articles/a-brief-history-of-the-cluster-randomized-trial-design/> (last checked March 2024).
11. Lindquist EF. *Statistical Analysis in Educational Research*. Boston, MA: Houghton-Mifflin, 1940. See <https://www.jameslindlibrary.org/lindquist-ef-1940/> (last checked March 2024).
12. Chalmers I, Matthews R, Glasziou P, Boutron I and Armitage P. Trial analysis by treatment allocated or by treatment received? Origins of 'the intention-to-treat principle' to reduce allocation bias: Part 1. *J R Soc Med* 2023; 116: 343–350. See <https://www.jameslindlibrary.org/articles/analysis-of-clinical-trial-by-treatment-allocated-or-by-treatment-received-applying-the-intention-to-treat-principle/> (last checked March 2024).
13. Chalmers I, Matthews R, Glasziou P, Boutron I and Armitage P. Trial analysis by treatment allocated or by treatment received? Origins of 'the intention-to-treat principle' to reduce allocation bias: Part 2. *J R Soc Med* 2023; 116: 386–394. See <https://www.jameslindlibrary.org/articles/analysis-of-clinical-trial-by-treatment-allocated-or-by-treatment-received-applying-the-intention-to-treat-principle/> (last checked March 2024).