Experimentalism in Dutch Education Policy:

Experiences and Lessons Learned

Essay by Lex Borghans, Trudie Schils and Inge de Wolf

November 2015

1 Introduction

Essay on experimental education research

In recent years, the Netherlands has accumulated a great deal of experience of experimentation in education. Over the last decade, there has been a shift in thinking about innovations in education, a shift in which experimental research has played an important role. The idea behind experimental research is that it results in a better understanding of what works in education. It has the potential to prevent situations in which educational reforms designed on the drawing board fail to have the intended effects on education once they are introduced. By systematically comparing different ways of organizing education, we can obtain a clearer picture of what works in practice, not only in the classroom but in the school as a whole. It is thought that, by taking this approach, we can improve education based on knowledge that has actually been tested in practice. A large number of education experiments have been set up in recent years, making use of a quasi-experimental approach to gain an impression of what works in education.

In this essay, we describe the Netherlands' recent experience of setting up experiments as part of its policy on education. Over the past decade we have learned many valuable lessons about how experimentation in education works and the problems we encounter in this area. We will chart these developments and describe the factors that have led to success or failure when conducting experimental research in education during this period. In doing so,

we will distinguish between the various phases of a project: the start of the experiment, the execution of the study and the dissemination of the results. Particular attention will be paid to the different perspectives of the various parties involved in experimental research in education, including teachers, school heads, governors, administrators and researchers. The essay also examines the implications of experimental research for the way in which policy is made: what lessons can be learned from our experiences of experimental research in education?

What are experiments in schools?

Experiments in schools are studies in which the effect of an approach (intervention) is examined by comparing a group of pupils who underwent the intervention (the experimental or intervention group) with a group of pupils who did not undergo the intervention (the control group). An intervention can take many forms, such as a new teaching strategy or a new package of educational materials, to name but two. A particular feature of experimental research is the random assignment of pupils to one of two groups. This random assignment means that both groups are comparable, as their make-up is based on chance, so differences in outcomes between the groups can be interpreted as the effect of the intervention. Many choices are made in the field of education. In doing so, everyone will choose options that reflect their individual situation. Relating outcomes to choices made by individuals or schools is therefore risky, leading to a situation whereby the forms of education selected by members of society who were always more likely to be high achievers will automatically be seen as better. Experimental research is intended to investigate causal effects. In science, randomized experiments are often seen as the 'gold standard' (a score of 5 on the Maryland Scientific Methods Scale).1It is used across a range of disciplines, most prominently in the medical sciences. For decades now, experimental research has also been used in other disciplines, such as psychology, economics, criminology, sociology and education sciences.

The randomized division of pupils into intervention and control groups is an important feature of experimental research because it offers a guarantee that both groups are similar in composition. This is of particular importance in education research, since many choices in

1Sherman et al. (1997) developed a scale for rating the methodological quality of scientific studies. Sherman, L. W., Gottfredson, D., Mackenzie, D., Eck, J., Reuter, P. and Bushway, S. (1997) Preventing Crime: What Works, What Doesn't, What's Promising. Report to the U.S. Congress. Washington, D.C. U.S. Dept. of Justice.

education result in pupils and students being immersed in a different educational experience and following alternative educational routes. So-called 'selection effects' are always present, and this makes groups of pupils difficult to compare. For example, to discover why some people spend longer in the education system than others, it is not enough to compare people who spend more time in the education system with those who spend less time (e.g. people who complete a university degree and those who enter vocational education). Such a comparison will produce a distorted view of the effects of education since the two groups being compared not only differ in the amount of education they receive, but can also be distinguished by pre-existing differences, such as their ability to learn. A simple comparison of the results achieved by both groups would therefore reflect not only the effects of additional education but also the difference in people's other characteristics. To effectively measure the effect of education on future outcomes, prospective university students would have to be assigned to vocational education and vice versa. Although often not feasible in practice, this principle is a benchmark for experimental education research. Since education is a very important part of people's lives, it is to be expected that people who have made different choices in education will also differ in other ways. Randomization is therefore more than a method of ensuring that the intervention and control group have the same composition. It is crucial to eliminate differences that almost certainly exist as a result of selection effects.

From cohorts to experiments in schools

In education, experimental research calls for a very different approach to the one used until recently in empirical education research. In education, it was customary to conduct studies using datasets containing longitudinal or other types of data intended to monitor the development of a group of pupils over time. Sometimes researchers collected the data they needed for this purpose on an ad hoc basis, but since the 1970s the Netherlands has also established a number of education cohorts, designed to study a large group of pupils over a long period, a process that involved testing and administering questionnaires to pupils, parents and teachers (COOL, SECP, PRIMA). This enables researchers who have access to this data to carry out analyses of education at their desks. Experimental research, however, requires a completely different way of working. In a number of schools one or more intervention and control groups have to be set up, with the intervention groups undergoing an adjustment in their education. This necessarily involves cooperation with schools, and thus with governors,

school heads, teachers and other stakeholders. Support for the experiment among all stakeholders is essential. The intervention being made must be practically applicable, and the pupils at the schools involved in the experiment should be monitored over an extended period. Newer initiatives in data collection within education therefore follow all pupils in a given region, while working more closely with the education sector in order to have a data infrastructure for evaluation when experiments are carried out (e.g. the Limburg Education Monitor). The use of experiments to help determine the direction of educational development seems to have great potential in terms of bringing about actual improvements in education. However, this does not mean that such an approach can be taken for granted or that it is bearing fruit.

A comparison with aerospace

When considering a new experimental approach and the changes in education research it implies, parallels can be drawn with the field of astronomy.2For a long time, astronomers relied solely on telescopes to study the planets and the stars. Just as an education researcher was able to conduct analyses at his desk, the astronomer did not have to step out from behind his telescope. The transition from traditional research to experimental research in education might be compared to the step from telescope-based research to space flight.

In 2007, around the time that this comparison was made, NASA launched the space probe Dawn. In 2011 Dawn flew past the asteroid Vesta and in March 2015 it reached the dwarf planet Ceres. Both of these bodies are thought to have been created at roughly the same time as the Earth, but they underwent a very different development. By comparing them, scientists are hoping to learn more about the origins of the Earth. Carrying out this type of scientific study is of course an entirely different proposition to studying space through a telescope. It requires incredible far-sightedness and meticulous planning. NASA began to prepare this mission as early as 1996 and an extraordinary amount of time was invested in the research itself. Although the probe was launched back in 2007, it has only just reached Ceres. The stakes are also high in terms of costs. In total the mission cost approximately half a billion dollars. These factors are part and parcel of experimental research. Compared to

2This comparison is also made in Borghans, ‘Kunnen we meer leren over leren?’, a lecture given before the Education and Labour Committee of the Social and Economic Council of the Netherlands (SER), The Hague, October 2007.

traditional research, experiments call for more extensive preparation and careful consideration has to be given to deciding which experiment can best be implemented. Because the effects of education may extend over a long arc of a person's career, sound experimental research requires a lengthy commitment. But the information that can be obtained is far more precise than can be obtained via traditional routes.

The aerospace sector was not built in a day. To successfully launch rockets and space probes and to carry out measurements across astounding distances, scientists began with small-scale test flights and learned from their failures. Even with all the experience we have accumulated, rockets still malfunction. Or success is only partial, as was the case with the Philae lander: it was lowered onto a speeding comet from the Rosetta space probe but it failed to latch onto the surface securely. In such cases, many years of hard work and dedication result in less information than was originally envisaged. Of course, a great deal can be learned from such setbacks, increasing the next mission's chances of success. This will also apply to experiments in education. It is not reasonable to expect that the introduction of an experimental working method will result directly in perfect interventions. This too will be a process of trial and error.

2 The importance of experiments

Learning by experimenting

Experimental research has long been common in science. At the beginning of last century, Frank B. Morrison described in *Feeds and Feeding* how farmers could divide their land into small sections to test different approaches on randomly selected areas. Trial and comparison is crucial to finding out what works better.

Experimentation is not unique to science. Teachers, for example, also experiment a lot. A study by Rivkin, Hanushek and Kain (2005) shows a sharp rise in the quality of a teacher's work during the first two years of his or her career.3 A likely explanation for this is that novice teachers discover through trial and error how to teach in an effective manner. They notice that what they are doing is not working satisfactorily, so they try a different approach for a few days and, depending on whether they think it is more effective, they switch to the

3 Rivkin, Hanushek & Kain (2005) Teachers, schools, and academic achievement. *Econometrica*, vol. 73, no. 2, 417-458.

new approach, which then becomes the basis for further experimentation. The ideas for a new approach may well come from colleagues who have drawn upon their own experiences. Every day teachers spend using an approach that later turns out to be relatively ineffective will reduce their productivity. As they increasingly discover what works, this will bring them closer to the ideal way of teaching.

Duration, size and cost of experiments

Not all improvements in education can come from teachers' own experiments. There are two important reasons for this. First, education is all about the pupils' development in the longer term. Most teachers see their pupils for only one year, making it difficult for them to relate the effects of their teaching to outcomes further down the line. In addition, a sufficient number of pupils is needed in order to achieve a meaningful comparison. The magnitude of the effect times the square root of the number of pupils determines the accuracy with which a comparison can be made. This means that a teacher is well able to observe large differences between approaches on the basis of a class of, for example, 30 pupils. For smaller effects – which, incidentally, can still be very substantial – it may be necessary to compare thousands of pupils with one another. A single teacher cannot achieve such a level of comparison, and a more systematic approach is needed.

This makes experimental research in education relatively expensive. Important effects are mostly longer-term effects and the conditions in education are difficult to control, so that the desired effect is often small compared with the variety of influences to which pupils are exposed. Richard R. Nelson and Sidney Winter cite the cost of research as an explanation for the development of science through time.4It is self-evident that researchers will start by analysing correlations for which only a small sample is required. Large effects with few environmental influences are, as it were, the low-hanging fruit that is picked first. That may explain why experimental research first emerged in agriculture, medicine and psychology. Furthermore, in the early stages of research in those areas you can see a particular focus on brief interventions with effects in the short term. In medicine, for example, drugs are tested experimentally, while much nutritional advice is still based on traditional longitudinal research. The benefits to be gained by research also play a role. It is interesting to note that agriculture was making use of experimental research as far back as the early 1900s. Farmers

4 Nelson, R. & Winter, S. G. (2009). An Evolutionary Theory of Economic Change, Harvard University Press.

have a strong commercial interest in good farming techniques. In all likelihood, the banks – as party that stood to gain from a farmer's success – also played an important role in encouraging experimental research. Education research is not only relatively difficult due to the long-term nature of expected effects and the strong influence of external factors, but also due to a lack of commercial pressure.

However, the importance of a good education is increasing. A great deal of research shows that the value of education in society is continuing to rise. The resources for giving young people more education over a longer period are continuing to decline, increasing the importance of making the years they spend in education as effective as possible. In addition, the cost of education research is continuing to decrease, most notably due to the emergence of IT. Many data relating to pupils, in particular test data, have already been entered into computer systems, and the introduction of a single registration number for people in education in the Netherlands (*onderwijsnummer*) is making it increasingly easy to link data to arrive at a bigger picture. On the one hand, it is therefore becoming more valuable to know what does and does not work in education and on the other hand it is becoming easier to monitor pupils over longer periods of time in education. Because the tracking of pupils is especially valuable if there are proper intervention and control groups to address crucial questions in education, this will only increase the value of experimentation.

3 The setting in the Netherlands

Scope for experimentation and innovation

Freedom of education is a fundamental principle of the Dutch education system, anchored in the nation's constitution at the beginning of the 20th century after a hard-fought political struggle. Freedom of education means that groups and movements within society are at liberty to establish their own schools and, as long as they meet certain conditions, that these schools are eligible for the same funding as public schools. At the time when this principle was adopted, most of the movements that established their own schools did so on religious grounds. Today the religious identity of a school is far less important to much of the Dutch population, but 'special-status schools' are still largely free to design and organize their education as they see fit. In the meantime, the government has been increasing the autonomy given to public schools, creating an overall setting in which schools and school boards have a

high degree of educational autonomy while the government generally exercises restraint when it comes to imposing regulations.

Nevertheless, there is often tension between this freedom of education and the government's desire to manage education and encourage improvement. This is perhaps best illustrated by a number of educational reforms implemented in the 1990s, which – in the perception of many – had an adverse effect on educational standards. These were the introduction of a stronger emphasis on independent study in senior years of secondary education (*studiehuis*) and the foundation cycle *(basisvorming*) in the early years of secondary education, and the merging of the basic tracks in general education and vocational education to form a single preparatory vocational secondary education track (VMBO). Other examples of government influence on education from that period include reduction of class sizes in primary schools, encouraging IT in primary education, and mergers and upscaling in vocational education. The public response to a number of these reforms was rather negative, giving rise to a general mood that the standard of education in the Netherlands was in rapid decline.

A parliamentary committee led by Jeroen Dijsselbloem was set up to look into these developments and to identify any lessons that might be drawn from them. One of the committee's conclusions was that, while it could not be said with any certainty that these reforms had led to a drop in the standard of education, it was safe to conclude that the government's control over the situation was not firm enough to prevent a loss of quality. With a view to organizing educational improvement more effectively in future to achieve genuinely positive effects, the committee highlighted the importance of gathering good data so that education can be monitored properly: existing evidence should be examined more closely before proceeding to implementation and experiments would be valuable instruments in this regard. The committee also concluded that the government had been interfering too much with educational reforms and that greater autonomy should be returned to the schools themselves.

In theory, the autonomy of schools and school boards in the Dutch setting creates optimum conditions for variety and scope for experimentation. Yet at the same time, the evidence suggests that these autonomous schools are not always inclined to learn from one other. They sometimes cooperate within the same administrative or collaborative setting, yet at the same time it is clear that dissemination of knowledge about what 'works in education' is not an automatic process. The widespread use of ready-made teaching methods and reliance on consultancy firms and organizations is another factor which does not contribute to mutual

knowledge-sharing between schools. As regards experimental education research, this is a missed opportunity, since it means that investment in an experiment at a particular school is unlikely to benefit other schools.

4. History of experimentation

International trend

From an international perspective, experimental education research has grown dramatically in recent decades. A number of large-scale experiments, such as the STAR experiment, have served as major examples. In the 1990s, particularly in the field of economics, people became keenly aware that research into the effects of certain forms of education that is based on a comparison between people who have made different choices leads to distortions in the results. This is because such choices are usually underpinned by particular reasons. A naive comparison between such groups amounts to a comparison between people who were already different in the first place and who have then gone on to receive a different form of education. The observed effect is therefore equal to the sum of the selection effect and the actual effect. The selection effect may be larger than the actual effect, and may also be either positive or negative. If people take a course without possessing the requisite knowledge to embark on it, the link between participation in the course and knowledge can be smaller than the actual effect of the course (or perhaps even negative). Research into reintegration programmes in the United States revealed the opposite effect. The more this study took selection effects into account, the smaller the observed effects were. This would seem to suggest positive selection. Experimental and quasi-experimental research (i.e. research in which coincidences that occur in the real world are used as a substitute for an actual experiment) began to be used increasingly as an alternative to the existing methods.

In the 1990s in particular, studies took a creative approach to excluding selection effects by making use of natural variations or stark contrasts within education policy (natural experiments and regression discontinuity studies). A classic example is Angrist and Lavy's study of the effects of class size5, which makes use of clear and unambiguous rules regarding allowable maximum numbers. This study initiated a whole tradition of natural experiments

5 Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2), 533-575

and quasi-experimental research designs. In this regard, an important role was played by the US National Bureau of Economic Research (NBER), which has published an extensive series of quasi-experimental working papers on education over the past decade and in doing so has established a major trend. Meanwhile, the leading international scientific journals also began to favour experimental and quasi-experimental studies over more traditional education research.

The problem with natural and quasi-experiments was that the most readily available natural experiments were utilized while other important questions about education went unanswered because no natural experiment was available to examine that question. This led to a shift towards establishing field experiments in order to answer specific questions. The World Bank has been a major initiator in this area. For some time now it has been financing mostly experimental studies on education in developing countries.6This has resulted in a series of field experiments in and around schools in countries such as India, Colombia and Kenya. In recent years, this line of research appears to be continuing, with educational experiments in schools in Western countries such as the United States. Major players in this area are the School of Education at Stanford (with Hanushek, Hoxby, Loeb & Bettinger), and Roland Fryer (The Education Innovation Laboratory).

The international growth of experimental education research has been triggered by a number of developments. Firstly, a number of governments have placed a strong emphasis on obtaining more knowledge about what works in education. For example, in the United States the proven effectiveness of an intervention is often a condition for obtaining funding. The World Bank has also actively contributed to more experimental research. In addition, particularly in the United States, investments have been made in the review and dissemination of the results of experimental education research. For instance, there are two major databases/websites which schools can use to identify effective programmes and interventions: the What Works Clearinghouse (WWC) and the Best Evidence in Education website (Robert Slavin's BEE). Reviews of proven effective teaching methods are also undertaken on a regular basis. By encouraging experimentation and dissemination of the results, a culture has emerged in which schools recognize the importance of primarily using programmes whose effectiveness has been proven. Parents may also play an active role in this process by asking their children's school to make use of demonstrably effective programmes.

6 Among others, see Banerjee, A.V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122 (3), 1235-1264.

Another important development is that the academic world has also embraced experimental and quasi- experimental education research and is also investing in this area. As indicated above, the leading journals now favour experimental studies over more traditional studies in education. Many prominent figures in education research invest in educational experiments, thereby increasing the number of good examples available. In the economics of education, we can now speak of a paradigm shift and other disciplines would appear to be following suit.

Hessel Oosterbeek and Scholar/TIER

In the Netherlands, the trend towards more experimental education research has followed these international developments. Professor Hessel Oosterbeek has played a particularly important part in raising awareness of the importance of experimental education research in the Dutch context. Here, too, the research is often quasi-experimental in nature since the opportunities for researchers to design experiments are limited. In his inaugural address in 2001, Professor Oosterbeek stressed the importance of research of this kind: *"To be certain that there is a causal effect when evaluating each* [education measure]*, the ideal approach would be to carry out a field experiment based on the gold standard of random allocation.* [...] *Unfortunately, that is not always possible, but fortunately there are alternatives. These alternatives are derived from field experiments, and are therefore referred to as quasi-experiments. Among economists, the term 'natural experiments' is more common. The strength of field experiments lies in random assignment. In quasi-experiments one looks for conditions that mimic this random assignment as closely as possible."* Professor Oosterbeek showed that the use of this method has led us to draw very different conclusions about certain educational matters since the 1980s. He pointed out that it takes time for policy-makers to fully realize the impact of this way of working, but that such a realization is important for the acceptance and success of the changes to be implemented in education. He also argued that too little attention is paid to evaluating the changes made. Changes born of political compromise and greeted with scepticism by the teaching profession are particularly unlikely to have their utility demonstrated at a later stage.

At the same time, Professor Oosterbeek also discussed movements in the opposite direction, whereby the then Minister of Education and State Secretary for Education advocated giving schools greater freedom to allocate their own resources. The theory was that

by giving sufficient freedom to schools in conjunction with the right incentives, schools would be able to realize their full potential. However, he also issued a warning in this regard: *"But those in the teaching profession know no more than the minister about how the ministerial budget should be spent [...]. Teachers cannot be sure that the method they are using, or would like to use, is better than another method. Perhaps the most obvious way to illustrate this is with the example of class sizes. When we see how much effort it takes researchers to show that reducing class sizes leads to better academic performance under certain conditions, it is utterly impossible for teachers themselves to possess this knowledge. In our opinion, therefore, a misguided comparison with market forces in the private sector is being made. In the private sector, companies that are unable to use the best available technology to deliver a product that meets the needs of consumers will go bankrupt. Surely we do not want to see the percentage of businesses that go bankrupt each year matched by the percentage of failed schools? Or parents and children being forced to look for another school on a regular basis? In this regard, it would be inadvisable to repeat New Zealand's experience of deregulation. Instead of the market analogy based on how private companies operate, it would be more appropriate to draw a parallel with the development of medicines. Patients and individual doctors in the field do not know whether a drug will work or not. And individual doctors in a local hospital are not in a position to find out. If education is ever to rise above the level of quackery, it is clear that we still have a long way to go. But it's clear in which direction the road is heading and it is also clear that this is the only right way to proceed."*

Dinand Webbink and the CPB

Around 2005, the Netherlands Bureau for Economic Policy Analysis (CPB) also began to place greater emphasis on the importance of experimental or quasi-experimental research. This shift took place under the leadership of Dinand Webbink. The CPB performed a number of quasi-experimental impact studies and in discussing international literature about what works in education, it placed increasing emphasis on experimental evidence. This is reflected, for example, in Canton and Webbink's 2004 survey of the effects of performance-related pay in education, in which they recommend experimenting with various forms of individual

performance-related pay for teachers as one of the policy options:[7] *"So far little is known about the effects of individual performance-related pay in education. The experiment conducted in Israel demonstrates convincingly that pupils' performance improves. This suggests that the introduction of individual performance-related pay in the Netherlands could be a promising prospect. However, within the current policy context, which is aimed at strengthening institutional autonomy, it is up to school managers to decide whether to make use of individual performance-related pay. Moreover, in recent years schools have started introducing competency-based remuneration and are currently gaining experience of this system. Educational institutions may well prefer to await the outcome of these experiences before starting down a new path with the introduction of individual performance-related pay. This means that the scope for central government to implement policy in this area is limited. Policy options include communicating the potential benefits of individual performance-related pay to educational institutions and incorporating performance-related pay into the current competency-based system. In addition, carefully designed experiments can provide more insight into the advantages and disadvantages of individual performance-related pay in education. Given that institutions are free to pursue their own remuneration policies, proceeding on the basis of voluntary participation in experiments would be the obvious choice."*

Alongside its own impact studies, the CPB redoubled its efforts to produce an overview of what it referred to as promising education policy (i.e. education policy shown by Dutch or international quasi-experimental research to be effective and efficient). The report entitled *Kansrijk kennisbeleid* (Promising Knowledge Policy) lists a large number of options. One of the CPB's tasks is to calculate the macroeconomic effects of the election manifestos published by the Netherlands' political parties. The effects of education policy form part of this process. While the CPB's approach has been an important stimulus for basing policy on knowledge about effectiveness, it gives rise to two problems. Firstly, it turns out to be difficult to translate experiments from other countries to the Dutch situation. A large number of environmental characteristics may influence the effectiveness of a certain intervention: a measure that works in one country, will not necessarily work elsewhere. In addition, experimental research is not readily available for all relevant policy decisions about education. It is very difficult to compare two policies if one has been the subject of research

7 Canton, E. & Webbink, D. (2004), Prestatieprikkels in het Nederlandse onderwijs: Wat kunnen we leren van recente buitenlandse ervaringen?. The Hague, CPB Document 49.

while the other has not. Basing education policy on experimental research to the greatest possible extent also means ensuring that research is conducted into crucial education questions (preferably within the Dutch context).

Other contributions: the Education Inspectorate and the Education Council

The Education Inspectorate is also interested in amassing greater knowledge about effectiveness in education. The Inspectorate has increased its reliance on data since the 1990s, especially as a supplement to school inspections. It uses data to identify schools at risk and to plan extra inspections accordingly. The Inspectorate also cooperates with experimental and quasi-experimental studies aimed at mapping out differences between schools and programmes. Along with the CPB, they also carry out experimental or quasi-experimental research into the effects of monitoring the learning achievements of primary school pupils.8In addition, the Inspectorate seeks to encourage schools to incorporate available knowledge about what works into their policies. One of its initiatives in this regard is contributing to the creation of a Dutch (TIER)website called Best Evidence in Education.9

In 2006, the Education Council also expressed its support for a more evidence-based way of working: where and how could an evidence-based approach contribute to the effectiveness of teaching? Experiments were seen as the keystone of this approach. The Council therefore made recommendations to encourage the uptake of evidence-based teaching methods in educational practice:

"A more evidence-based approach requires a culture change. Knowledge of the effectiveness (or ineffectiveness) of teaching methods and approaches should be more accessible. A digital service desk should be established to make research data on what works more accessible by means of reviews and other sources. Partly on the basis of evidence-based insights, educational institutions should make more conscious decisions when choosing certain objectives and methods. However, schools are and remain free to choose."

Traditionally, the Netherlands has a fairly extensive network of agencies that support schools in implementing innovation in their teaching. In addition to national educational

8 Luginbuhl, R., Webbink, D., & de Wolf, I. (2009). Do inspections improve primary school performance? Educational Evaluation and Policy Analysis, 31(3), 221-237.

9 See www.bestevidenceineducation.nl

centres, such as SLO and CITO, there is a regional network of support agencies that were initially funded directly from public resources. Later these funds were made available to the schools, so that they could purchase support as they saw fit. In two subsequent reports10, the Education Council found that the approach taken by these agencies appears to bear little relation to scientific evidence. The Council argues that schools and the agencies that support them should make far greater use of the findings from research. But it is also important that we learn more about how education works. To this end, schools should join forces with one another and with research institutions to work on data collection and meaningful experiments. Schools which undertake these activities, are also contributing to the future of the sector as a whole. It is therefore evident that funds should be made available for this purpose. The Education Council observes that in the Netherlands very little is spent on R&D in the field of education. Coalitions of schools and knowledge institutions should be given the opportunity to obtain funding for research that addresses jointly formulated questions. The main criterion for the approval of any such plans should be sound and thorough design. The content should be left to the educational and research institutions themselves.

The Education Council therefore suggests a strategy whereby educational experiments are not set up by the government on a large scale, but whereby experimental design is left up to the schools and school boards. This proposed approach respects the principle of freedom of education and implies that – even with the use of the experimental method – educational reform is not dependent on centralized control, but is the result of a creative process in which schools themselves, in mutual cooperation, determine the direction of innovation.

A change in thinking at the Ministry of Education

Partly due to these developments, the Dutch Ministry of Education is also giving increasingly serious consideration to the importance of experiments in the improvement of education. In 2007, it set up a knowledge division with responsibility for a large number of initiatives relating to experimental education research. In 2009 a major research programme called *OnderwijsBewijs* (EducationEvidence) was launched, a programme in which schools can join forces with knowledge institutions to request education experiments on a number of themes. In 2010 the Ministry launched *Zicht op effectiviteit* (With a View to Effectiveness) and

10 Education Council (2010), Ontwikkeling en ondersteuning van onderwijs en onderwijsraad (2011), Ruim baan voor stapsgewijze verbeteringen

provided funding for the setting up of the TIER institute, whose aim is to generate evidence-based knowledge about education. 2011 saw the launch of the *Innovatieimpuls* (Innovation Impulse) programme and the decision to experiment with performance-related pay in education. In 2012, the National Directorate for Educational Research (NRO) was established and in 2014 legislation was introduced requiring schools to implement an anti-bullying policy based on a demonstrably effective method.

The funding of education research in the Netherlands operated along a number of different channels. The ministry allocated resources to educational support organizations and organizations covered by the National Education Support Activities (Subsidies) Act, the Netherlands Organization for Scientific Research (NWO) whose Programme Council for Education Research (PROO) dealt with education research and funded research programmes directly. To combine these resources and to focus more on the questions that matter most to those working in education, the National Directorate for Educational Research (NRO) was established. Due to the government's depleted financial situation, the NRO's budget turned out to be lower than initially anticipated. Experimental and evidence based research is not a major asset of NRO. Its main asset is enable is to narrow the gap between schools and academic research. NRO organizes meetings with the teaching profession on relevant themes and enables the teaching profession to initiate research. Whether or not these initiatives are experiments or evidence based research, is of less importance. A second complication is that setting up experiments with schools is a complex process, far more complex than other types of educational research. Intensive cooperation between schools and academics and time are crucial to set up and conduct experiments. The NRO-process, with a central determination of themes, bureaucratic procedures and short time between calls and submission dates hinders experimental research. There is little scope for educational institutions and researchers to enter into dialogue and to design experimental research proposals reflecting issues that are important to teaching professionals. A last complicating factor is that experimental research is often more expensive than other types of research. A major part of the cost of experimental research lies in the adjustments that have to be made to the educational process in the intervention schools. The resources needed to bring these adjustments about often exceed standard research budgets. These three complications (no asset, the NRO-processes and budget restrictions) make that few large-scale experiments get off the ground within the NRO-programme.

5 Overview of experiments

Experiments within the *OnderwijsBewijs* programme[11]

In 2009, with support from the Ministry of Finance, the Ministry of Education launched a research programme called *OnderwijsBewijs* (EducationEvidence), which enabled schools and knowledge institutions to apply jointly to take part in educational experiments on a number of themes. The programme consisted of two rounds. In the first, 18 grants were awarded for the themes of giftedness, language learning and arithmetic, teacher shortages, continuous learning pathways, early childhood education and child welfare. In the second round, 19 grants were awarded for the themes of behavioural problems and bullying, reducing backlogs, excellence and citizenship.

These included both projects initiated by a strong impulse from the teaching profession and projects initiated by the world of research. The experiences from the first round showed that while experiments in education are a wonderful idea in theory, putting them into practice is anything but straightforward. It proved particularly difficult to involve sufficient numbers of schools in university-initiated projects. Level of participation was less of a problem in projects where the research question came from a group of schools, whether working in combination with a university or not. Projects initiated by an individual school tended to encounter problems with the experimental design and face issues of generalizability and scalability. The most successful projects were those created when a group of schools experiencing particular educational problems enlisted the help of a university to formulate the research question and to design an experiment to determine the most effective approach to the problem. In some instances, misunderstandings arose about the design aspects of experimental research. For instance, some of those involved turned out not to be aware of what a randomized trial was. Interestingly, though a number of project groups were convinced that randomization simply was not possible in their particular case, randomization ultimately proved possible in almost all of the projects and achievable by means of a design that was acceptable to those involved.

11A more detailed description of the design and experiences can be found in: Wolf, I. & Borghans L. (2012) Ervaringen OnderwijsBewijs. In: Waterreus, I., Van der Heul, I. (eds) Stapsgewijze verbeteringen in het onderwijs en samenwerking tussen onderwijsonderzoek en onderwijspraktijk. Pedagogische Studien, 89, 377-387.

For an experiment to provide a clear picture of whether an intervention works, it is essential that a sufficient number of pupils/schools participate. In light of this, it is striking that in a large number of projects given the go ahead, the sample size was rather small. Research plans often turned out not to be based on an analysis of statistical power that indicates how large the group of participants should be. Another challenging aspect of experimental design is the protocol used for the control group. Some projects excluded pupils in the control group from every aspect of intervention as much as possible. Amid the complexities of real-world education, comparison with current practice is often easier and more interesting to explore. For example, this enables us to test the impact of a new method or additional teacher-pupil interaction. Comparing this with the traditional method shows the additional effect of the new method. However, this does make it more difficult to prove additional effects and the implications for statistical power also need to be thought through. Since the expected size of the effect is smaller, it means that the research often has to be carried out on a larger group of pupils, classes or schools. Experiments in which the control group continues to use the traditional method meet with fewer objections. In such a design, the members of the control group are not denied anything; their exposure to the new method is only postponed until such times as its effectiveness has been tested.

Some experiments started with a pilot project. Problems with the actual execution of experiments can be largely overcome by conducting a pilot project in the first phase of a study. This involves trying out the intervention in a small number of schools. A pilot of this kind is of great value, not least in providing a so-called process evaluation for the experiment. What are you likely to encounter when carrying out the experiment in practice? What solutions are available? There are always issues which neither the teacher/school head nor the researcher have anticipated but which can have a major impact on the execution and results of the experiment. A pilot can prevent disappointment due to teething troubles and may even provide information about the expected effect size, information which can then be utilized in the design of the experiment. Many of the above-mentioned experiments set up from within the teaching profession are in fact closer to pilot projects than full-blown experiments because they are essentially geared towards “trying something out”.

Last but not least, it is important to monitor the execution of the experiments. The assumption that an intervention has been made in accordance with the researchers’ stipulations often proves to be erroneous. Coordination with and the cooperation of schools, teachers and pupils is crucial to implementation. Selective drop-out in the control group is

often the biggest source of concern, often caused by a lack of understanding about the intervention.

Experiments and quasi-experiments by the CPB

In recent years, the CPB has occasionally been involved in evaluations in education. Where possible it has attempted to take an experimental or quasi-experimental approach. These studies also produced interesting experiences with the experimental or quasi-experimental method. It turned out to be very difficult to get schools to carry out experiments. For this reason, the CPB used quasi-experiments which sometimes involved asking the teaching profession to follow procedures that increased evaluability. This tendency to look for quasi-experimental opportunities rather than implementing complete experiments created tension between the questions that one would prefer to have answered and the questions it was possible to answer with the data available.

Sometimes it turned out to be very difficult to emulate a good control and intervention group with the available data. This can be illustrated by a study of the effectiveness of additional supervision and support for underperforming schools in Amsterdam.12Since the municipality applied this intervention to all weak schools in Amsterdam, it was not possible to find a control group within the municipality. The study therefore turned to other municipalities. However, this too proved problematic since it required making a comparable selection of schools in other municipalities, based on the assumption that these municipalities were not pursuing other policies that might have an effect on school performance. The more such assumptions have to be made, the more the quasi-experimental method ends up resembling traditional research.

A special case is the study of community schools (*wijkscholen*) in Rotterdam, the effectiveness of which was also evaluated by the CPB. The community school is an initiative whereby pupils who are in danger of falling through the cracks in the system can receive an education to improve their job opportunities or guide them towards another educational programme. Because the initiators of this approach are very much against the random allocation of places at the community school – their philosophy is that every pupil is entitled to use this facility – the CPB decided to make use of the fact that there are only limited places

12 Van Elk, R. & Kok, S. (2014), Het effect van de Kwaliteitsaanpak Basisonderwijs Amsterdam op leerlingprestaties: Resultaten van de eerste vier jaar. The Hague, CPB Discussion Paper no. 264.

available and that in some cases community schools have to turn pupils away simply because they are full. The control group therefore consisted of pupils who were referred to the community school yet were unable to attend because there were no places available at the time. The question of why this form of selection was seen as less objectionable than the randomized system that was rejected in the first place will remain unanswered here.

Since the ministry was keen to gain an insight into the effects of the community school as quickly as possible, the CPB produced an interim report.13At the time of the interim report, a significant proportion of the pupils were still enrolled at the community school. That made a comparison with the control group difficult. After all, many of the pupils in the control group were no longer in education. When pupils still attending the community school were included in the analysis of how many subjects were in employment or training, the results were bound to show a favourable effect for the intervention group. And if these pupils were excluded from the analysis, it would be difficult to identify a relevant comparison group, as it is not known which pupils in the control group would have still been attending the community school if there had been a place for them. The CPB conducted numerous robustness analyses and despite these problems came to the conclusion that the community school had a positive effect on the careers of the young people who attended it.

Some time later, the final evaluation followed. This was easier to perform because by that time almost all of the pupils had left the community school.14Remarkably, this evaluation showed no positive correlation between attending the community school and transition to education and employment. While this reversal in findings was no doubt a painful confrontation for those involved in the community schools project, it does provide strong evidence for the usefulness of the experimental approach: a thorough analysis that appeared to come close to replicating an experimental study, nevertheless produced very different results.

In addition to the effects on the transition to education and employment, the final evaluation also looked at the impact of the community school on crime. The study showed that those who attended the community school were in fact more likely to get into trouble with the police. A breakdown of the results into pupils who had been in trouble with the police before attending the community school and those who had no police record showed that the increase in criminal behaviour only applied to pupils who had previously been in trouble with the police. For the other groups the career effects of attending the community

13 Van Elk,R. (2011), Evaluatie wijkscholen Rotterdam, The Hague CPB Memorandum.
14 Van Elk, R., van der Steeg, M. & Webbink, D. (2013), De effecten van de wijkschool in Rotterdam op onderwijsdeelname, werk en criminaliteit. The Hague, CPB Discussion Paper no. 241.

school were shown to be beneficial, however, although the observed effects were not significant. That could mean that for some pupils the community school leads to more criminal behaviour, while other pupils experience beneficial effects with no negative effects in terms of criminal behaviour. If the study had been more extensive, or if it had been based on a random assignment of pupils, for example, these favourable outcomes may well have been significant. This shows that seemingly minor details in the design of a study may have greater effects on the reported findings.

Under the heading *Zicht op Effectiviteit* (With a View to Effectiveness), the Ministry of Education commissioned the CPB and Ecorys to come up with experimental or quasi-experimental designs for the evaluation of policies.15In a first round, designs were made for all of the ministry's policy areas. In a number of cases it proved impossible to design an experiment that satisfied the so-called gold standard of experimental research. A number of other designs were actually carried out, but no follow-up took place aimed at finding solutions for those policy evaluations that did not seem to fit the existing mould.

Experiments with performance-related pay

In 2011, a new government was formed: a coalition between the liberal VVD and the Christian democrat CDA. One policy measure in their coalition agreement was that resources should be made available to introduce performance-related pay for teachers. Performance-related pay was one of the promising educational improvements recommended in the reports of the CPB. Partly because the Ministry of Education was keen to ensure a support base within the teaching profession and partly because it was not clear what form of performance-related pay would be most effective, the government decided to initiate this process by carrying out a number of experiments. Schools were invited to submit proposals for performance-related pay which, if they resulted in an adequate impact assessment, would be subsidized by the Ministry of Education. Since groups of schools or school boards were being given the freedom to come up with their own interpretation of a performance-related pay programme, it serves as a prime example of how experimentation and a relatively high degree

15 Van Elk, R., van der Meer, F., van der Steeg, M. & Webbink, D. (2011). Zicht op effectiviteit van beleid: Studie naar evaluatieontwerpen voor onderwijs- en wetenschapmaatregelen. The Hague, CPB Background Document.
Briene, MFM and the Vlasakker, S. (2011). Zicht op effectiviteit beleid: evaluatieontwerpen beleidsinterventies cultuurdomein. Rotterdam Ecorys.

of school autonomy could go hand in hand. By testing the various interventions in different groups of schools, it would be possible to see what worked and what did not. In a traditional intervention only a single implementation of the planned adjustment can be tried out. A conceptual problem associated with this diverse approach was that the intended experiments would only show the effect of a proposed form of performance-related pay within the schools that opted for that particular form. In other words, if an approach were found to be effective, it would not necessarily mean that the same effects would occur at other schools which adopted the approach. Strictly speaking, this could only be established by means of a randomized follow-up experiment.

Two problems arose in the run-up to these experiments with performance-related pay. Firstly, many of the performance-related pay proposals developed by the schools bore little or no relation to what was known about this subject from the scientific literature. Secondly, in a number of cases the contact between the researchers and the schools involved was far from ideal. The approach that some researchers took to the design of an experiment often turned out to be far removed from the realities of school life. A gulf existed between the researchers' ideas of how an experiment should be conducted and the experimental possibilities that exist within a school setting.

In fact, this represented a collision between two contrasting visions of what constitutes science. On the one hand there was a vision of science as a particular prescription that must be followed in order to produce good research: analyses that follow this prescription are scientific, those that deviate from it are not. This was the view held by a number of researchers and also by the Ministry of Education. For example, the ministry wrote that use should preferably be made of the 'gold standard' with respect to experimental research, but that compromises should be made where necessary.

Deviating from the standard prescription is therefore regarded as less scientific. However, it is also possible to see science as the attempt to establish a systematic way of trying to establish the existence of certain effects as effectively as possible, given the actual conditions. Once again, the comparison with aerospace research can be drawn. It is far simpler, for example, to carry out soil analysis on Earth than to do so on a distant planet by sending a space probe. If the analyses on Earth are more accurate than the measurements on the distant planet, this does not mean the latter are not scientific. The scientific challenge lies in developing methods whereby the problems encountered while gathering data on a distant planet are alleviated as much as possible.

This is similar to the challenge we face with regard to experimental education research. We are familiar with the ideal of a randomized experiment and this forms an attractive prospect for education research. The scientific challenge is to set up experiments in a school context which benefit as much as possible from the power of the experimental approach while coming up with solutions to any problems that arise along the way.

In the end the coalition government was short-lived and the experiments with performance-related pay did not materialize. This was partly because performance-related pay was a highly sensitive issue for the trade unions, one which provoked fierce union opposition. They were convinced that performance-related pay was not feasible and would diminish rather than enhance the motivation of teachers. It might be argued that the doubts surrounding the effectiveness of this instrument made it an ideal candidate for experimentation, but given the alternative logic that governs political processes, it would probably have been better not to initiate the experiments in the first place.

Experimental anti-bullying programmes

A subsequent step in the promotion of education on the basis of proven effectiveness was taken in the form of legislation governing anti-bullying policies in schools. In recent years, there has been extensive coverage of the negative effects of bullying in schools, resulting in a political and social consensus that schools should take action to combat bullying. Legislation was drafted requiring schools to have an anti-bullying policy based on a programme whose effectiveness has been proven by research. A committee was set up to assess whether anti-bullying policies met the requirements. Due to contradictions with the freedom of education principle, proposals for a mandatory effectiveness test were withdrawn. Perhaps solutions to the issue of bullying are too closely bound up with the identity of schools in the Netherlands, most of which are religiously oriented, to permit an approach with such a mandatory component. Mandatory effectiveness assessment may yet become a quality requirement with regard to other themes, for instance as a way of preventing the sale of all kinds of teaching methods to schools without their effectiveness having been clearly established. The experience in relation to this anti-bullying legislation may prove useful in this regard, as coming up with a sound research assessment procedure is no trifling matter. Firstly, such a system should provide ample scope for the effectiveness of untested methods to be evaluated

experimentally. Secondly, the anti-bullying issue has generated a good deal of debate about the evaluation criteria to be used.

Other experiments

Experimental education research faces another problem: certain data is required in order to carry out experimental or quasi-experimental analysis, yet it is simply not possible to generate or tap into the appropriate data for every interesting question that arises from an educational perspective. This is nicely illustrated by the doctoral research conducted by Ferry Haan. As a journalist and teacher, Ferry Haan has a strong commitment to education and is currently working on a PhD thesis under the supervision of Professor Hessel Oosterbeek. His study is an attempt to answer questions that he considers important in education and to analyse them in a rigorous manner. This has turned out to be perfectly possible for some questions, while others remain unanswered. Research into the effectiveness of 'Steve Jobs schools' (schools with a strong focus on modern IT) and IMC weekend schools turned out to be impossible because the problem of selection bias turned out to be insurmountable. No schools of this kind offered quasi-experimental opportunities in the shape of a surplus of applicants, an admission policy based on a lottery or something of that kind. However, such an opportunity was presented by the summer schools pilot project, which provides extra tuition during the summer months to enable pupils to obtain a pass in subjects they failed first time round to avoid having to repeat a year at school. The aim was to make use of the application surplus for summer schools for the purposes of evaluation. Schools volunteered in dribs and drabs but with the summer holidays fast approaching the researchers started with the first schools that had agreed to participate. In the end there were enough participating schools to justify a study, but because places were not allocated randomly, the schools that volunteered early took part while those that volunteered later did not. This led to a form of selection, which meant that the research was no longer possible. It is interesting to note that while many politicians were quick to praise the summer schools as a success, in fact their performance has not yet been the subject of a proper evaluation. The fact that many pupils who attended the summer schools progressed to the next year was seen as a success but to date we have no way of knowing how many pupils would have achieved the same result without attending the summer school.

Another plan was to analyse the innovations in the teaching of economics at secondary school level. These innovations involved providing more context, which in turn required

pupils to read more text as part of their programme. The plan was to examine the effects of this change on the differences in performance between boys and girls. However, this turned out not to be the only innovation taking place. As the pilot projects were getting under way, a number of subdivisions within school subjects were abolished, making it impossible to analyse the effects of the content-related innovation in its own right. The one subject which did provide scope for comparison was mathematics. A study was carried out which showed that changes in the teaching of Mathematics B at secondary school (HAVO) resulted in more girls opting for Mathematics A.

In addition to the above-mentioned evaluation with regard to mathematics, it has also proved possible to evaluate a number of other projects. One of these is a regression-discontinuity analysis of excellence programmes at three schools. At one school, the study has been completed and the results are both large and significant. Another analysis is focusing on data from two schools which randomly tested half of their first-year pupils on intelligence, eagerness to learn and their responses to the school questionnaire. The research is focused on whether this will improve the level they achieve beyond first year.

6 Lessons, dilemmas and opportunities

Interventions and comparability

It is often thought that experimental research cannot make a fair comparison unless the pupils in the control group and the pupils in the intervention group meet exactly the same conditions. This idea is often used to argue that experiments cannot be conducted in an educational setting. After all, pupils are continuously open to all kinds of influences outside the school environment. However, the good thing about randomized experiments is precisely that they ensure that outside influences do not affect the validity of the experiment. Since pupils are assigned to one of the two groups on a random basis, pupils affected by other influences in addition to the intended intervention will be divided approximately equally across intervention and control group. This means that distortion in the comparison between the two groups is avoided. However, the greater the variety of influences that pupils undergo, the smaller the effect of the intervention (measured in terms of relative influence), and the larger the group of pupils required to measure this effect with the same precision. Since the precision depends on the square root of the number of pupils, four times as many pupils are required if there are twice as many differences in environmental influences. While the

experimental method is particularly effective in detecting the impact of a specific intervention even if there are many other influences at work, researchers often try to limit these influences as much as possible, so that they can keep the sample size small.

Ethical concerns

Ethical concerns are frequently cited as reasons not to carry out experimental education research. The argument is that it is unethical to provide some pupils with an intervention while withholding it from others. Especially when it is assumed that an intervention will be highly effective, withholding it from a control group is used as an argument for not using the experimental method to test the effectiveness of the intervention. Of course, it is doubtful whether this conclusion can be reached so simply. If the effect of the intervention is indeed beyond any doubt, it is worth asking why an experiment is needed at all; in most cases it is uncertain whether or not an approach will have the desired effect. Why do such objections not apply in the medical sciences, for example, where medicine is tested experimentally? And why would exposing pupils to an untested intervention be more ethical than testing the effectiveness of an intervention? What if it transpires that the intervention is not effective or may even have an adverse effect? We have the impression that the ethical argument is often used as a readily available argument against experimental research, when in fact there are underlying objections on other grounds. In projects where all parties are more involved in the design of an experiment, the ethical argument is less likely to surface.

With regard to this particular aspect, experimental education research can learn a great deal from the medical sciences, where researchers always seek to identify any ethical concerns beforehand. When such concerns are found, solutions are often sought and submitted to an ethics committee. In experimental education research, this practice is far less common. However, we recommend that it should be adopted, especially when it comes to issues of privacy and enriching research data.

Interestingly, there is sometimes tension between the value of an experiment for a participating individual (pupil, teacher) and its value to society (education, schools). Experimental research can yield a great deal for education as a whole, but for the pupils and teachers who participate, the research is sometimes of less benefit. The intervention is usually withheld from the control group, for instance, which means that the direct value to members

of that group can sometimes be disappointing. However, there are ways to address these concerns, such as offering a postponed intervention once the results of the experiment are known (if the intervention proves effective).

Duration, size and cost of experiments

Experimental education research is not always cheap. The cost of small-scale experiments, for example, within a class, is often reasonable. But when the research involves multiple classes and multiple schools, the costs quickly rise. The design and implementation of interventions in educational practice and the measurement of the effects on pupils is especially time-consuming. The drain on financial resources can rise dramatically as a result, especially when the experiments become larger or more complex. It is worth pointing out that in their design and when measuring results, studies do not always make use of existing resources such as measurement data already compiled by the school and pupils' registration data. Yet the use of such data is relatively inexpensive and leads to a huge increase in the usefulness and feasibility of experimental education research. For example, it becomes relatively easy to study the added effects of an intervention, such as effects on pupil performance, school careers and other long-term effects.

The magnitude of an effect is seen by statisticians as the correlation between the effect of the intervention studied and the extent of other influences on the development of the pupil. As a study covers a longer period of time, other influences become larger. Long-term research therefore requires a larger number of participants than short-term research.

Bringing together expertise in experimental research

These experiences of experiments in education bring three types of expertise to the fore. All three are required to set up and carry out an experiment effectively:

(1) Knowledge of the education sector: this is knowledge of how educational practice operates and is organized, in the classroom and the school as a whole.

(2) Scientific/content-related knowledge: this is knowledge of the theoretical and empirical literature in a specific area.

(3) Statistical knowledge: this is knowledge of the design of an experiment, how randomization works and how problems affecting randomization can be overcome.

Some of the problems setting up experiments in education arise because a single party rarely possesses these three types of expertise, and there are often differences between how a school and how a researcher approach an experiment. Ensuring that the three different types of relevant expertise come together in a research team can result in an experiment that is workable in the educational setting, the effects of which are attributable to the intervention and the results of which are of both scientific and practical value. Progress has been achieved in this area in recent years. For instance, an increasing number of consortia are being formed between researchers and professionals from the educational setting when it comes to the design and implementation of an experiment. However, at present such consortia are only temporary in nature, a response to the requirements of a given subsidy programme. More sustainable solutions to the experimentation problems outlined above can be provided by long-term partnerships between schools and researchers, giving rise to an ongoing dialogue about the problems that schools experience and what scientists can provide in terms of literature, whereby those involved get to know and understand each other's world.

Schools participating in experiments

There is a difference between experiments that are initiated on the basis of policy or research and experiments that are the result of an impetus from within the teaching profession. Policy-makers and politicians are often driven by the need to answer questions about whether a particular policy measure is working or not. The problem being studied or the intervention being implemented is not necessarily shared by all schools. How would you then set about determining the participation of schools in such an experiment? There are at least two possible ways to approach this. One way is to determine the participation of schools in research randomly, for example by means of a lottery procedure organized by the Ministry of Education (aside from the matter of which participating schools then become control or intervention schools).

Another approach would be to open up the research to all relevant schools and invite applications to participate. The latter approach best reflects a situation in which schools have the freedom to organize their own education. In that case any effect found will be relevant to those schools alone, as schools that voluntarily participate in the research do not constitute a

representative sample of the total population of schools. It could be that some schools are eager to participate in experimental research because they are struggling with the specific problem being addressed in the study. Additionally, it may be that schools object to participating in a study in which the division into intervention and control groups is determined by random allocation, with no assurance of being part of the intervention group. School heads who decide to participate in a study may differ in the extent to which they are prepared to accept such uncertainty and this may affect participation in experimental research. This prompts an expectation that the findings would say something about how the tested measure/policy would pan out if it were to be applied to all schools. However, a disadvantage is that the tested measure/policy need not be useful to all types of schools and thus it would not be useful to involve all schools in the study.

Relationship to data collection

The development of experimental research cannot be viewed separately from the availability of data. Since education is crucial to personal development, the important outcomes of educational interventions only really become visible in the long term. The Netherlands has a long tradition of data collection in the context of education cohorts: since the 1970s, regular studies have taken place which monitor a large group of pupils as they pass through primary and secondary education. For a long time the primary and secondary school cohorts were conducted separately but with the introduction of the COOL cohort study, an attempt is being made to monitor pupils through both primary and secondary education.

Many educational experiments make no use of such longitudinal data. This means that, in many cases, only the short-term outcomes of specific interventions are examined. Establishing links with the ongoing cohort studies would be difficult, because this implies that the schools participating in the cohort and the schools participating in the experiment (as intervention or control school) would have to be the same.

Statistics Netherlands is increasingly using administrative details as part of its data collection. The organization has adopted this approach so that all sources of data on individuals can in principle be related to one another by means of the citizen's service number. This will gradually result in a large longitudinal file that covers the entire population. This will ultimately do away with the problem of the group of schools where an experiment is

carried out having to be identical to the cohort group. Additionally, this administrative approach enables pupils to be monitored for a very long time.

At present, the administrative data collection amassed by Statistics Netherlands mainly contains details of people's employment situation. In recent years, however, more and more data on education has become available. Some of the above-mentioned cohort studies have been linked to Statistics Netherlands' administrative data, which provides data on the further life experience of pupils who were originally only monitored during their years at secondary school. For the further development of experimental research it would be of great value if, for the schools that form part of the study, data about the further career of pupils were to be made available through this route.

Dissemination of results

A final important lesson to be learned from the range of educational experiments in the Netherlands is that improvements need to be made regarding the dissemination of results. The large degree of autonomy enjoyed by Dutch schools and the lack of knowledge exchange between them means that relatively few schools benefit from the results of experiments. In this regard, too, bridging the gap between teaching practice and scientific knowledge is no mean feat. Given the cost of experiments and the valuable insights they produce, this is a great pity. There are a number of initiatives geared towards sharing knowledge about 'what works' with professionals from educational practice. For instance, TIER has developed and launched a Best Evidence in Education website16, there is a website which features the best practices and results of experiments conducted as part of the *OnderwijsBewijs* programme17 and Manzano and has written several books on interventions in education which have been tested and found to be effective. These are all sympathetic initiatives, but not widely used within the teaching profession. In this respect, there are clear differences between the situation in the Netherlands and that in the United States, for example, where much greater emphasis is placed on the dissemination of the results of experimental studies (e.g. through What Works Clearinghouse, BEE, incentives for stimulating effective methods). Moreover, here in the Netherlands it is often the government or the academic world that takes the initiative and encourages the dissemination of measures or good examples.

16 See www.tierweb.nl/bee/tier-bee.html

17www.onderwijsbewijs.nl

7 Conclusion

Experimental research pays

Over the last decade, the Netherlands has amassed a great deal of experience of experimentation in education research. Especially in combination with the development of good data collection where pupils' long-term development can be monitored over the longer term, this experience has great potential as regards achieving systematic improvements in education.

The randomized experiment is elegant and simple, yet carrying it out in practice is far from straightforward. We have a long way to go before the question of how best to design experimental research that is effective in educational practice can be fully answered, and it will require much more in the way of scientific creativity to develop intelligent approaches for this purpose. Nor is it the case that well-designed experimental research always leads to irrefutable answers. Details in the design of a study may influence the outcome, as with any type of research. Although policy-makers need clear answers, the power of science must continue to reside in the fact that all findings remain open to discussion. Even conclusions and interpretations that are almost universally accepted at a given point of time can be seen in a new light as a result of new research or new approaches.

Challenges for the future

To ensure the success of experimental education research, constructive cooperation with the teaching profession is crucial. At present, this is a major bottleneck in the development of experimental education research. In an ideal world, schools and school boards would test any changes they plan to make using an experimental approach, before proceeding to implementation. Given that such changes in education are often projects that go far beyond standard research budgets, effective coordination between researchers and the plans within the educational setting is of the essence. At present, this aspect is still fraught with difficulty and there is often a lack of cooperation and dialogue between research, policy and the profession. Organizing such cooperation is also a complex matter. But if our aim is to use experiments to answer bigger questions and look at long-term effects, cooperation must amount to more than collaboration on a one-off experiment in one or several school classes. Such cooperation is not easy to organize; it requires investment, perseverance, scope for experimentation and

good mutual relationships. As yet, such partnerships and relationships between teachers and schools, scientists and policy-makers are few and far between. However, there are a number of fledgling partnerships in the Netherlands, and this gives us cause for hope. Examples from the United States show that this form of cooperation is not only possible, but also leads to important new insights for education.

A second challenge concerns the selection of themes and interventions to be examined. Experimental research provides insight into the effects of interventions, but does not answer the question of what research needs to be carried out for the further improvement of education. Nevertheless, important choices need to be made in this respect. The matter of how to make such choices and how to facilitate experimental inquiry into important questions that need to be addressed represent a major challenge for the development of experimental education research. If we expect our schools to base their policies on what is known to work and if we assess the plans of political parties on what is known about their effectiveness, it is important that the knowledge they need in order to take such decisions continues to be generated. At present, choices with regard to experimentation are too often dependent on available data, policy themes or ad hoc questions from schools. As a result, some experiments are not designed as well as they should be, some important experiments are never carried out at all and other experiments focus on interventions that have no prospect of ever being implemented by teachers or schools. When drawing up a solid research agenda for experimentation, it is advisable to seek out or bring about cooperation between the teaching profession, the academic world and the policy-makers. It is also important to base choices on knowledge about effects in education, both national and international.

A third and final challenge is to organize the learning ability within teaching. An important step in this direction would be to improve the dissemination of the results of experimental education research. This could reduce reluctance among schools to take action and make them less dependent on incidental choices from the world of educational advisors. An important precondition is that teachers and schools should be given access to scientific publications and other sources of research data. Investments can also be made in review studies and websites that increase the accessibility of results for schools, following the example set by countries such as the United States. Yet greater dissemination alone is not enough. It would also be highly beneficial if schools themselves were to experiment more and go in search of research partners for this purpose. This process could be more readily facilitated, not only in time and money but also by means of legislative scope. Policy-makers can also do much more to ensure that schools mainly use effective programmes, for example

by expecting this of them or by providing incentives to do so. Parents and pupils/students can also make demands in this regard. In addition, education advisors can perhaps play a more active role when it comes to stimulating the learning ability of schools. In the long term, it would be wonderful if schools and the academic world joined forces to take responsibility for the tradition of experimental education research, doing away with the need for government involvement altogether.