The long history of Direct Instruction research: Part 1

Kerry Hempenstall

We hear a lot these days in education about the importance of evidence-based practice and explicit teaching. Any word search in current educational policies will produce numerous exhortations to educators to make use of these features in their curricula.

For example, announcing the *Review to Achieve Educational Excellence in Australian Schools* (2017), the document proclaims "The Australian Government is committed to evidence-based reform", and evidence-based appears three times in its two pages. In the report of the *Teacher Education Ministerial Advisory Group (2014)*, 'evidence-based' appears 31 times, and one recommendation was that: "The theory, methods and practices taught to pre-service teachers need to be clearly based on evidence linked to impact on student learning outcomes" (p.18).

In the report of *Senate Standing Committee on Education, Employment and Workplace Relations (2013)*, the term explicit occurred 10 times, and in the National Inquiry into the Teaching of Literacy (2005) – 40 times!

What is less well known is that an approach to teaching known as Direct Instruction (DI) has arguably the largest evidence-base of any current model of instruction; the earliest programs having been developed in the 1960s and many evaluations have occurred since that time. Further, it is the source from which explicit instruction emerged.

> *" … it is clear that the roots of explicit instruction come directly from Direct Instruction and direct instruction, both of which have a history of effectiveness, especially for students with, and at-risk for, LD."* (*Hughes, Morris, Therrien, & Benson, 2017*, p.145)

So, to possibly confuse the issue, along with Direct Instruction there is also a model known as direct instruction (lower case). This latter term was initially introduced by Bereiter and Engelmann in their 1966 publication, *Teaching Disadvantaged Children in the Preschool*.

> *"The direct-instruction approach assures that every objective can at least be attended to and it gives the teacher better day-to-day control over pupil progress so that she will know what objectives need additional attention."* (p. 56)

Around 1968, Engelmann and colleagues coined the upper case term: Direct Instruction, when they began employing the acronym DISTAR (Direct Instructional Systems for Teaching and Remediation) to identify their programs.

Lower case direct instruction became more broadly known when Barak Rosenshine and David Berliner first employed it in 1978. Along with others, such as Evertson, Brophy, Good, and Stevens, their work during the 1970s-1990s on process-product research (examining what teachers do in the classroom and relating these teaching behaviours to student outcomes) established what became known as the effective teaching movement. The associated effective behaviours became known as direct instruction. The two definitions are certainly related, and

upper case DI programs incorporate the principles enunciated in the lower case di research as described below. However, as we shall later see, DI also provides the curriculum content in addition to the delivery system of direct instruction. Notice in the definition below, direct instruction (which later morphed into explicit teaching) indicates that the teacher must choose or provide the curriculum. That is, curriculum design and content are not part of direct instruction.

> *"Direct instruction [di] pertains to a set of teaching behaviours focused on academic matters where goals are clear to students; time allocated for instruction is sufficient and continuous; content coverage is extensive; student performance is monitored; questions are at a low cognitive level and produce many correct responses; and feedback to students is immediate and academically oriented. In direct instruction, the teacher controls the instructional goals, chooses material appropriate for the student's ability level, and paces the instructional episode."* (Rosenshine & Berliner, 1978, p. 7)

This approach of replicating the procedures used by effective teachers (those whose students had superior outcomes to those students of other teachers) was demonstrated to be valuable in a range of correlational and then experimental studies, such as by Good and Grouws (1979) in mathematics, and Anderson, Evertson, and Brophy (1979) in reading.

The evidence base of Direct Instruction

A 2018 paper published in the *Review of Educational Research* outlines and analyses the long history of research into the effectiveness of the various Direct Instruction programs: The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research, and its results may surprise those who have been inclined to dismiss it as an instructional option.

> *"Quantitative mixed models were used to examine literature published from 1966 through 2016 on the effectiveness of Direct Instruction. Analyses were based on 328 studies involving 413 study designs and almost 4000 effects. Results are reported for the total set and subareas regarding reading, math, language, spelling, and multiple or other academic subjects; ability measures; affective outcomes; teacher and parent views; and single-subject designs. All of the estimated effects were positive and all were statistically significant except results from metaregressions involving affective outcomes. Characteristics of the publications, methodology, and sample were not systematically related to effect estimates. Effects showed little decline during maintenance, and effects for academic subjects were greater when students had more exposure to the programs. Estimated effects were educationally significant, moderate to large when using the traditional psychological benchmarks, and similar in magnitude to effect sizes that reflect performance gaps between more and less advantaged students."* (Stockard, Wood, Coughlin, & Khoury, 2018, p.1)

> *"The strong positive results were similar across the 50 years of data; in articles, dissertations, and gray literature; across different types of research designs, assessments, outcome measures, and methods of calculating effects; across different types of samples and locales, student poverty status, race-ethnicity, at-risk status, and grade; across subjects and programs; after the intervention ceased; with researchers or teachers delivering the intervention; with experimental or usual*

comparison programs; and when other analytic methods, a broader sample, or other control variables were used." (Stockard, Wood, Coughlin, & Khoury, 2018, p.22*)*

These outcomes are impressive given the wide range of study designs, sample sizes, educational domains, and evaluation tools employed across the studies. Although there were variations across programs, effect size for the total sample was 0.60, with the 95% confidence interval within 0.54 to 0.66. This is a little lower than previous meta-analyses that analysed smaller samples, such as *White's 1988 meta-analysis* (25 studies in special education) which reported an effect size of 0.84. In the *Adams and Engelmann* meta-analysis in 1996, 37 research articles met the criteria for inclusion, producing an effect size of 0. 87. More recently, John Hattie (2009) reached broadly similar conclusions about the size of effect:

"One of the common criticisms is that Direct Instruction works with very low-level or specific skills, and with lower ability and the youngest students. These are not the findings from the meta-analyses. The effects of Direct Instruction are similar for regular (d=0.99), and special education and lower ability students (d=0.86), higher for reading (d=0.89) than for mathematics (d=0.50), similar for the more low-level word attack (d=0.64) and also for high-level comprehension (d=0.54), and similar for elementary and high school students. The messages of these meta-analyses on Direct Instruction underline the power of stating the learning intentions and success criteria, and then engaging students in moving towards these. The teacher needs to invite the students to learn, provide much deliberative practice and modeling, and provide appropriate feedback and multiple opportunities to learn. Students need opportunities for independent practice, and then there need to be opportunities to learn the skill or knowledge implicit in the learning intention in contexts other than those directly taught." (*Hattie, 2009*, pp. 206-7)

"One of the common criticisms is that Direct Instruction works with very low-level or specific skills, and with lower ability and the youngest students. These are not the findings from the meta-analyses."

For greater detail of evaluations into each of the various programs, see Writings on Direct Instruction: A Bibliography.

An important element in the 2018 meta-analysis is the durability of effects. It is a well-known issue in program evaluation that published programs may be shown sometimes to display a worthwhile effect immediately following intervention, but either no follow-up is instituted, say in six months or a year, or if it is performed, the effects appear to have washed out over that period. This often occurs with short-term interventions, and in those in which insufficient feedback and practice are incorporated.

Other important finding is that of a dose-response relationship, that is, the effects become larger if students are provided with more exposure to the programs. This counteracts the potential explanation of success that any apparent effects in the short term are due to novelty – the increased motivation wrought by participating in a new program. Unsurprisingly, the programs proved more powerful when introduced early in students' school careers.

"Earlier literature had led us to expect that effect sizes would be larger when students had greater exposure to the programs, and this hypothesis was supported for most of the analyses involving academic subjects. Significantly stronger results appeared for the total group, reading, math, and spelling for students who began the programs in kindergarten; for the total group and reading for students who had more years of intervention; and for math students with more daily exposure. Although we had expected that effects could be lower at maintenance than immediately post-intervention, the decline was significant in only two of the analyses (math and language) and not substantial in either." (Stockard, Wood, Coughlin, & Khoury, 2018, p. 22-23)

For findings of other reports and studies on DI, see Reviews supporting Direct Instruction program effectiveness Updated 2018.

Some argue that small studies and those with a variety of designs are inappropriate inclusions in a meta-analysis. It is obviously important to examine the highest quality research – experimental studies with random allocation, because they provide good internal validity. That is, they provide a measure of confidence that any effects noted can be attributed to the intervention, rather than to extraneous variables. Small quasi-experimental studies can be flawed in various ways; however, error is diffused and less of concern when consistent effects are noted across many studies (*Stanovich & Stanovich, 2003*). So, we should not dismiss small studies or those with less sophisticated design. They can add balance, providing external validity that is often missing from small or short term randomised controlled trials.

" … observational data sometimes meet the assumptions of a quasi-experimental design, at least approximately, such that causal conclusions are credible. If so, the estimates of quasi-experimental designs – which

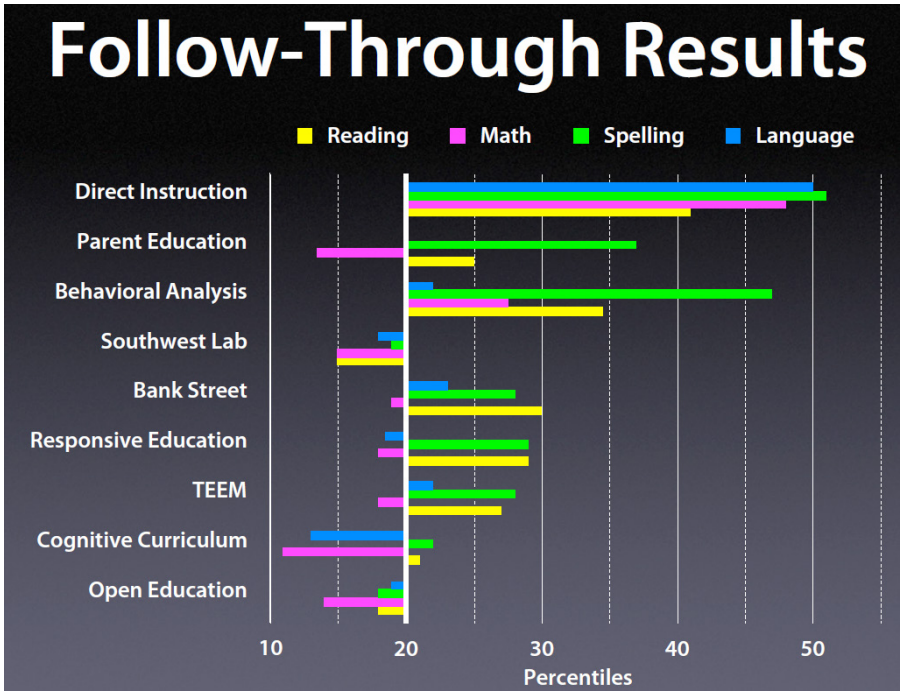
exploit naturally occurring selection processes and real-world implementations of the treatment – are frequently better generalizable than the results from a controlled laboratory experiment. Thus, if external validity is a major concern, the results of randomized experiments should always be complemented by findings from valid quasi-experiments." (Kim & Steiner, 2016, p.404)

A confluence of findings from numerous studies allows some confidence that the interventions will produce effects across a range of settings, not solely in the single experimental setting. So, the aggregation of data from many different studies is capable of producing a meaningful and valid conclusion (Slavin, 2003). However, that does not mean that studies with faulty designs should be included. The 2018 study's selection criteria led to the rejection of 221 studies for a variety of reasons, including insufficient information and methodological shortcomings. Of the 549 studies identified, 328 were subsequently included in the analyses.

"The over-arching evaluative concept educational practitioners should hold is that replicability of findings is the most important scientific standard for research findings to meet. That is, replicability of findings is the most useful form of evidence-based information of effectiveness, not the findings of a single study, no matter how well such studies are designed. In emphasising replicability, the logical structure of multiple-baseline designs (see Sidman, 1960) is a far more appropriate design framework for the evaluation of the effectiveness of instructional interventions than traditional group designs because they involve intrastudy replications of the effects of experimental interventions across what Campbell and Stanley (1963) call "time series". (Vitale & Kaniuka, 2012, p. 28-29)

Project Follow Through

The newer research analyses are welcome, but the findings are not new. Similar findings were reported in a huge study in the USA many years ago. This study was federally funded in the USA in the late 1960's, arising because of a concern about the poor educational outcomes achieved for disadvantaged students. Entitled Project Follow Through (Engelmann, Becker, Carnine, & Gersten, 1988), the study involved 75,000 children in 180 communities over the first three years of their school career. This was the largest educational experiment ever undertaken, extending from 1967 to 1995, at a cost of almost a billion dollars. There were comparisons across 20 competing sponsors covering a broad range of educational philosophies. They included models of child-directed learning, individualised instruction, language experience, learning styles, self-esteem development, cognitive emphasis, parent-based teaching, Direct Instruction, and behavioural teaching. The models can be reduced to three distinct themes – those whose instruction emphasised either basic academic outcomes, cognitive development, or affective development. The targeted basic skills for the evaluations were reading, language, spelling, writing, and maths. As it did in each of the other basic skills areas, in reading, the Direct Instruction model, which has a strong phonics emphasis, had the most impressive results in both academic and affective areas.



Source: Slocum, Stenhoff, and Van Schaack (2003)

"The Direct Instruction model had an unequivocally higher average effect on scores in the basic skills domain than did any other model. Finding 3: Where models have put their primary emphasis elsewhere than on the basic skills, the children they served have tended to score lower on tests of these skills than they would have done without Follow Through. All models other than those labelled "Basic Skills" had more negative than positive outcomes on measures in the basic skill domain." (Watkins, 1997, p. 32-33)

Follow-up studies were performed three, six, and nine years after the DI students completed Follow Through. They showed strong consistent long term benefits in reading (Gersten, Keating, & Becker, 1988); effects that were evidenced in higher achievement, fewer grade retentions, and more university acceptances than in comparison groups that had traditional education in the same communities.

> *"We offer Project Follow Through results as support for a direct, explicit approach to teaching; however, it is important to point out that although Direct Instruction includes the majority of the elements of explicit instruction and is based on such principles as increasing on-task behaviors, high levels of success, and content coverage, it is distinguished from explicit instruction by its emphasis on curriculum design (Stein, Carnine, & Dixon, 1998). Aside from this curriculum based distinction, the overlap of teaching procedures is extensive."* (Archer & Hughes, 2011, p.14-15)

What also may not be well known is the long history of DI research in Australia, particularly through Alex Maggs' contributions. The early studies include those by Becker, Engelmann, Carnine, and Maggs (1979), Booth (1978), Bracey, Maggs, and Morath (1975), Calder (1982), Clunies-Ross (1990), Fields (1986), Gersten and Maggs (1982), Kenny (1980), Leach & Siddall (1990), Lockyer and Maggs (1982), McLean & Moore (1985), Maggs and Moore (1978), Maggs and Morath (1976), Maggs and White (1982), Maggs (1976), Maggs and Moore (1983), Maggs and Murdoch (1979), Maggs, Moore, and Boldie (1978), and Taylor, de Lacey, and Nurcombe (1974).

What features of the model are most significant?

> *"The sponsors of the Direct Instruction model … developed the most effective instructional method that is currently available. They could not have done so, however, had they not looked at teaching as a technology and at learning as an orderly process. It is this view of learning that is critical to convey to the educational community. Educators must be taught that learning is a function of the student-teacher interaction, the instructional moment. They must learn that there are qualitative variations in those interactions and that the function of educational research is to determine what types of interactions, or methods, lead to the most change with the least resources."* (Watkins, 1997, p. 90-91).

For more reading on Follow Through, see Direct Instruction and Project Follow Through: A Bibliography.

So, DI has been around a long time – how many other approaches can you think of that have accrued a large body of supportive evidence over 50 years? OK, but surely that means it's old, and has been surpassed by other new and shiny approaches that take into account more up-to-date program development and research. That view may have some justification if the currently available DI programs were 50 years old. However, they have been constantly updated as new relevant evidence accrues. Further, the vast research data banks on each program are analysed by the designers to find details within the program structure or content that would benefit from re-writing in a new edition. Far from being moribund, there have been six new DI programs published since 2000. In the 2018 meta-analysis, more than half the research was conducted in the last 20 years.

In a second part to this paper, the major elements underpinning the Direct Instruction model will be described.

Kerry Hempenstall is a teacher and educational psychologist and Senior Industry Fellow, School of Education, RMIT University, Melbourne, Australia. In his work with the Victorian Education Department and RMIT University, he developed a particular interest in children's academic and behavioural issues. Email: shiraz1@iprimus.com.au