Assessing pupil development and education in an inclusive setting

Sip J. Pijl* and Dorien Hamstra

Department of Orthopedagogy, University of Groningen, the Netherlands

In the evaluation of an experimental, full inclusion model in the Netherlands, data on participating pupils with special needs were collected. The evaluation focused on social/emotional functioning, pupil behaviour and the quality of education. Independent assessors were invited to study the files of 24 pupils and evaluate their development and education. The assessors regarded the development of seven pupils as worrying. The remaining 17 were average or above average. The findings were compared with parents’ and teachers’ assessments. The discussion addresses the difficulties involved in interpreting evaluation results for Special Educational Needs (SEN) pupils in inclusive settings.

Introduction

The standard educational provision for pupils with special needs has long been a school or class for full-time special education. Nowadays, provisions for these pupils are more varied and much more oriented towards regular education (Pijl et al., 1997; Vislie, 2003). Many European countries strive for full-inclusion models with various provisions in regular education; others offer provisions in both regular and special schools; and only a few still have special schools as standard provision (Meijer, 1998).

This development towards inclusion is partly based on a critical analysis of the role of segregated special education (Clark et al., 1999), the long-term effects of special education placement (Tøssebro & Haug, 1998) and further on the belief that children with special needs benefit from inclusion (Cole et al., 1991). After all, support for inclusion would diminish quickly if it were found to have a detrimental effect on pupils. However, research has not yet found convincing evidence that inclusion is better — or, at least, no worse — for pupils with special needs than segregated special education (Nakken & Pijl, 2002; Pijl et al., 2003).

From an inclusive perspective, data on the effects of inclusion when compared with segregated special education have limited value (Booth, 1996; Söder & Pijl, 1996).

*Corresponding author: Department of Orthopedagogy, University of Groningen, Grote Rozenstraat 38, NL-9712 TJ Groningen, the Netherlands. Email: s.j.pijl@ppsw.rug.nl

ISSN 1360–3116 (print)/ISSN 1464–5173 (online)/05/020181–12
© 2005 Taylor & Francis Ltd
DOI: 10.1080/1360311042000331192
Hall (1996), for example, states that research on the effects of inclusion serves no purpose, as the fundamental issue is that all children have a right to attend a regular school. A more practical argument here is that countries with full-inclusion models have no special settings with which to compare. In these countries, the effect question is reformulated in terms of the quality of education for all (Eklindh, 2001).

However, in countries with provisions in both regular and special education and certainly in countries making the first steps towards inclusion, the effect question is often posed by researchers, politicians, teacher and parent organizations, and society at large. This question does not depart from the idea that any eventual negative effects should bring to a halt the inclusion process, but more from a monitoring and adjusting perspective, or what Dyson (1999, p. 42) describes as ‘the pragmatics discourse within the inclusive education movement’. Negative effects should lead to early stage adjustments of the inclusion process, thereby preventing further negative effects to both individual pupils and the inclusion process.

Research on the effects of inclusion faces several methodological problems (Farrell, 2000; Nakken & Pijl, 2002). There are problems in matching the pupils in regular and special education, difficulties in describing the differences and similarities of the two settings in such studies, i.e. integration and segregation, and controlling other factors influencing study results like educational programmes and teacher qualifications.

Yet, in a local, experimental, full-inclusion model in the Netherlands, evaluation data on pupils with special needs participating in an inclusive setting were required. To start this experiment, the Ministry of Education had to adjust regulations, make extra funding available and was accountable for the effects of the experiment. Positive effects not only would justify the Ministry’s support for the experiment, but also would support the budding inclusion processes in the Netherlands. Negative reports, on the other hand, would cause various problems, both within the experimental setting and for the new policy on inclusion. The findings of the evaluation were therefore of great interest to several parties involved both in the experiment and in the implementation of the inclusion policy in the Netherlands. The research question was twofold: how to measure effects on pupil level and what effects were attained in a local, experimental full-inclusion model in the Netherlands.

The setting

The education system in the Netherlands consists of both a regular and a separate special school system. In the past five decades, the special education system has shown a steady growth, resulting in 4.3% of all pupils aged 6–11 years being placed in different special school types in 1997 (Pijl, 1997). In 1996, a newly written policy paper (Ministry of Education, Culture and Science, 1996) suggested changing the organization of special schools drastically, reviewing the statement procedures and overhauling funding regulations in order to support inclusion special needs pupils in regular schools. This policy has become known as the ‘Backpack policy’: pupils take
funding with them to the school of their choice. Parliament has endorsed new legis­
lation and this policy is now fully operational.

Currently, there are four types of special schools, also known as expertise centres: 
those for visually impaired pupils, those for pupils with hearing and/or speech disor-
ders, those for physically and mentally disabled pupils, and those for pupils with
behavioural disorders.

During the preparation period for the new Backpack policy, the city of Almere
wanted to establish special schools. Almere, a new, growing town founded in 1976 in
the polders not far from Amsterdam, had no special schools for pupils with sensory,
motor and mental impairments and/or behaviour disorders. These children had to
travel to surrounding cities, which resulted in long working days and in limited social
contact in their own environment.

In the early years, the number of inhabitants in Almere was too small to start special
schools, but today the city has over 150 000 inhabitants and the city council hoped
to offer special needs pupils schooling in the city. The government regarded Almere’s
request as fair and understandable, but not in line with current developments in
educational policy. It decided to cancel plans for new special schools, but allowed
Almere to implement the newly developed policy plans on inclusion on an experi­
mental basis. While the whole country was still debating the 1996 inclusion policy and its
implications, Almere started implementing it in 1997.

Almere’s main goal was to ‘include pupils in regular primary schools in their own
neighbourhood’. It started in May 1997 with an ambitious target: to include 50% of
all pupils with special needs in the city within 5 years. This experiment covers 10 years
and is limited to primary school pupils. To offer the best quality of help, schools are
supported by an urban expertise centre. The centre’s so-called case managers are
responsible for coordinating all support to each special needs pupil. An average case
manager could have a caseload of five to ten pupils.

The expertise centre has developed three different kinds of inclusion models:

- Pupil in Group: the pupil is fully included and receives the same education as all
  the other children in his/her class. Sometimes the pupil follows (for some subjects)
a specially adapted curriculum. The groups all have a regular teacher. In 2002,
over 270 pupils were included in this model.
- Group in School: a group of pupils is placed in the pupils’ own classroom inside a
  regular school. An important goal is that all children should participate as much as
  possible with the activities of the rest of the pupils of the school. The group has a
special teacher and, often, a teaching assistant. Currently, there are three groups
comprising about 30 pupils.
- Group Coupled to School: a group of severely impaired children is placed in a
special class loosely attached to a regular school. In daily practice, this model best
resembles a school for special education. Where possible, the pupils in the group
participate in activities of the regular school. In the case of Almere, this group
consists of approximately 40 children with severe behavioural disorders (see
www.gewooanders.nl).
Method

From the foregoing, it is clear that in terms of models, organization, funding, experimental basis and scale, the project in Almere is rather unique in the Netherlands. The evaluation study of this project can be regarded as a case study with pupils as ‘embedded units of analysis’ (Yin, 1989).

Considering the methodological problems involved in comparing pupils in regular and special education settings, i.e. matching, differences between the education in regular and special settings, and controlling other factors (see above), and the often difficult to interpret results (Söder, 1997), it was decided not to work with a design involving a control group in special education settings. Another option was to compare the development of the pupils in Almere with special needs pupils placed in regular education settings elsewhere in the Netherlands. These pupils are fully integrated on an individual basis in regular schools and supported by a peripatetic teacher from a special school. However, matching these to pupils in Almere would only partially work because regulations limit the categories of special needs in peripatetic teaching, and research has shown that only relatively well-performing pupils are accepted in peripatetic teaching (Claessens et al., 1989). This would result in a number of special needs pupils in Almere without a match partner elsewhere.

Parents and regular school teachers of the Special Educational Needs (SEN) pupils in Almere are closely following the development of these pupils and have a say on their development. Their opinions are valuable in evaluating the inclusion process, but it is difficult for them to interpret their assessment in terms of which development is reasonable and expected for a particular child. After all, most lack data on the development of comparable children in other settings. The only professionals with much experience in supporting special needs pupils and their teachers in regular education settings are the peripatetic teachers working in regular settings. The limitations in participating in peripatetic teaching described above do set boundaries for the pupils with which they are working in regular education, but most peripatetic teachers also have much experience in special education. They are therefore the best-informed assessors available. The developments of the SEN pupils in Almere will be assessed by asking both parents and teachers in Almere, as well as independent assessors outside the city, to evaluate pupils’ dossiers.

In 1998–99, two groups of pupils (cohorts 98–99 and 99–00) were followed for 3 and 2 years, respectively. The cohorts were randomly selected from the newly admitted pupils (71 in 98–99, 82 in 99–00) in the various inclusion models in Almere. In each year, 34 pupils/parents were invited to participate in the study, but of these 68, only 45 actually decided to take part. During the 3-year study, the sample lost another 14 pupils due to various reasons (transfer to another school, no longer categorized as having special needs, moving house). Of the remaining 31, six attended the ‘Group Coupled to School’ model. Since the research question for this paper focuses on the effects on pupil level in a full-inclusion model, the pupils in this group are excluded from the data reported herein.
The possible effects on pupil level were elaborated as social–emotional functioning, pupil behaviour, social integration, learning performances and the quality of the education of the pupil. The degree of social integration of the pupils was assessed by using sociometric techniques. In collecting data on learning performances, it became clear that little was documented in the schools. This resulted in often obsolete and incomplete datasets. In addition, the datasets for the sociometric analyses proved to be incomplete. It was therefore decided not to use learning performance and social integration as effect variables in this study. The social–emotional functioning, pupil behaviour and the quality of the education of the pupil will be the focus points.

The available data on pupils in the ‘Pupil in Group’ and ‘Group in School’ models consist of assessment, classroom observation, and questionnaire data based on interviews with parents and teachers. Based on these data, pupil dossiers were made. The dossiers presented data on the pupil’s background, described pupil behaviour at school, addressed the social–emotional functioning of the pupil involved, and described the education (curriculum and practice) of the pupil. A typical pupil dossier comprises five to seven pages of text. It starts with a description of the pupil’s home situation (parents’ ages, training and professions, information on siblings, and available data on how supportive, stimulating, warm or caring is the family situation). In almost all cases, extensive information on the pupil’s general development from birth is present. Reports on the medical history and early psychological, pedagogical and other assessment data are briefly presented. A large part of the dossier focuses on a description of the actual behaviour of the pupil in class and his/her social–emotional development. This is based on teachers’ and parents’ information, on data gathered during classroom observation and — where available — on recent test data. It covers behaviour at school, reactions to other pupils, disturbances in class as well as self-image, self-confidence, perceived competence and social skills. The dossier also comprises information on the education of the pupil, based on interviews with the class teacher, the case manager and head teacher. It addresses, among others, the school’s education goals, key aspects of the way of working in and outside the class, class organization and individual or small-group support.

The second author transformed all the available raw data on the pupil into a dossier. To control for possible researcher effects, a second, independent researcher was asked to read the files of three randomly selected pupils and to make a second dossier on these pupils. This was performed by a colleague also working in the field of special needs education, but in another department of the university and not in any way involved in the study reported herein.

In total 28 (25 + 3) pupil dossiers were drawn up based on 25 pupils. Eight addressed pupils with various communication impairments (deafness, hardness of hearing, speech/language problems), 14 were described as having motor and/or intellectual impairments, and the remaining three had severe behavioural disorders. Eighteen pupil were included using the ‘Pupil in Group’ model and the remaining seven were attached to a group in a school.

The dossiers were sent to independent assessors. Because it was regarded unlikely that they had specialist knowledge and expertise on the whole range of impairments,
three groups of assessors were formed: one for communication disorders, one for motor and/or intellectual impairments, and one for behavioural disorders. The assessors were randomly selected, but anyone working in regions around the city was exempt. Most assessors were peripatetic teachers with experience in supporting special needs pupils and their teachers in regular education settings. Further, a school psychologist and a special education director also participated. Nine assessors were involved: three for communication disorders, four for motor and/or intellectual impairments, and two for behavioural disorders. Each dossier was assessed independently by two assessors. Each assessor received a maximum of seven pupil dossiers.

The assessors were asked to read carefully the dossiers of the pupils and give a judgement on social–emotional functioning, pupil behaviour and the quality of the education of the pupil. These three assessments consisted of scores on a five-point scale, ranging from 1: very negative and dissatisfying situation, to 5: very positive and satisfying situation.

Gower's coefficient (Gower, 1971) was used as an index for agreement between assessments. It uses the absolute sum of differences between the assessments and is — when compared with Cohen's kappa (Popping, 1983) — less influenced by asymmetrical distributions. Gower's coefficient is 0 if there is no agreement at all and 1 if there is perfect agreement between assessments. The assessors were stimulated to give comments and explain their assessments. Most assessors provided extra information.

The assessors’ evaluations were compared with qualitative data based on interviews with the parents and teachers of pupils involved in the study (Hamstra, 2003). The pupils’ parents and teachers were asked to evaluate the pupils’ social–emotional development, developments in their behaviour and the adequacy of the education offered. These were scored as negative, neutral or positive. The three evaluations were combined into one score representing the parents’ and one score representing the teachers’ assessment. Four cases had missing parent and/or teacher data, resulting in 21 complete cases for this part of the analyses.

**Results**

**Adequacy of pupil dossiers**

Based on the files of three randomly selected pupils, both the second author and a colleague from another university department each made three dossiers on three pupils. The six dossiers were presented to two assessors. Agreement between the assessors was calculated using Gower’s coefficient, resulting in, respectively, 0.79, 0.87 and 0.70 for each of the three domains. This indicates that both the project leader and the independent colleague compiled comparable descriptions of social–emotional functioning, pupil behaviour and education of the pupil. The comparability of the latter category is somewhat disappointing, but the findings suggest that project leader effects in preparing the dossiers are small. Note that the project leader was a bit more positive in describing the quality of education compared with the report prepared by the independent colleague.
Reliability of assessments

To make sure that the various assessors would give similar assessments on social–emotional functioning, pupil behaviour and the education of the pupil, each dossier was assessed by two different assessors. A high agreement between the assessments shows that different assessors came to similar conclusions. Table 1 presents the data based on 24 pupil dossiers; one assessor never returned the assessments on one pupil. The findings in Table 1 address pupil behaviour, social–emotional functioning and the education of the pupil. Both the first column and the first row present the values of the five-point scale used by the assessors. The remaining 25 cells show the frequency of the various combinations of assessments. Frequency ‘1’ in the matrix on social–emotional development in row 2, column 4, indicates that in only one case does an assessor scores a ‘2’ (rather negative), while the second assessor was much more positive (4). The frequencies on the diagonal show the concurring assessments.

Gower’s coefficients on the data in Table 1 are 0.85, 0.78 and 0.72, respectively. That suggests that the assessors agree on the social–emotional development and on behaviour, but the agreement on the quality and effects of education is just sufficient. The conclusion is that different assessors come to comparable assessments if they applied the five-point scale on the pupil descriptions given. The assessments could therefore be used when evaluating the effects on pupil level.

Evaluating effects on pupil level

Table 1 already gives a first impression of the results of the evaluation. The intersections of the first two columns and rows in each of the three matrices show how many pupils were judged as being at risk. These represent combinations of the low scores (1, 2), i.e. ‘very negative and dissatisfying’ and ‘negative and dissatisfying’. Other combinations of scores, e.g. 1 and 3, 1 and 4 or 2 and 3, are not taken as a clear indication that the development of pupils or the quality of their education was (very) disappointing.

In the case of the combination of low scores, the assessors felt that the social–emotional development of five pupils was worrying, that the same holds true for the behaviour of two pupils and two pupils were not receiving a proper education and/or progressing as expected. A further analysis of the data shows there is overlap between the pupils referred to here (Table 2). Seven pupils were actually involved here: one pupil’s development in both the social–emotional and behaviour domains, and one pupil’s situation in both the social–emotional domain and education were regarded as worrying. These seven pupils represent 29% of the group being assessed. The assessors regarded the development of the other 17 pupils in this study as being average or above average level. The intersections of the last two columns and rows in each of the three matrices show how many pupils were assessed as functioning above average. Corrected for overlap, the social–emotional functioning, pupil behaviour and the education (quality and effects) of 14 pupils (58%) were assessed as being above average.
An analysis on data based on interviews with parents and teachers of the pupils described above (see above) shows that both parents and teachers of these pupils are much more positive about the pupils’ development and education compared with the independent assessors. Of the 21 parents involved, only two evaluated their child’s development and education as negative, and three were indecisive. A similar pattern holds for the teachers’ judgements: two negative and five indecisive. The large majority of parents and teachers came to a positive evaluation. In all but one case, negative or indecisive judgements of parents and/or teachers matched the negative evaluations of the independent assessors. The agreement between the independent assessors and parents and between the independent assessors and the teachers was calculated using Gower’s coefficient, resulting in an agreement of 0.64 and 0.78, respectively.

The further analysis focuses on the seven pupils assessed as having a (very) disappointing development and/or quality of education. Six of these pupils were individually integrated in the ‘Pupil in Group’ model, leaving one for the ‘Group in School’ model. This does not seem an unusual division, since three-quarters of the 24 pupils involved in the study were integrated individually in the ‘Pupil in Group’ model (see above).

The data in Tables 1 and 2 also show that the social–emotional development of the pupils has been a major issue for the assessors. In five of the seven cases, the
social–emotional development of the pupils was evaluated as a (very) negative and dissatisfying situation. In these five cases, the assessors’ comments clearly show that they worried about the (long-term) effects of both the family and school situation for these pupils. More specifically, they feared that staying in a classroom with more able pupils, having few or no contacts and relations, being teased and having conflicts could result in uncertainty, fear of failure, negative attributions, negative self-perception, etc. In three of these cases, the assessors suggested that the regular school underestimated the seriousness of the problems and doubted the decision to continue the inclusion process for these pupils.

In the case of pupil F (Table 2), the assessors also expressed their concern on the education of the pupil. In this particular situation, the school leader was extremely negative regarding inclusion and the classroom teachers faced a negative attitude, a lack of support and problems in getting appropriate materials. Finally, pupil G tended to isolate herself from the group and from the teacher. A reason could be that she became aware of the differences in abilities compared with the other pupils in the classroom and was beginning to lock out this frustrating world.

### Summary and discussion

This study shows that a group of nine independent assessors evaluated the situation of seven of 24 pupils as disappointing and worrying. Opposite to this 29%, 71% developed and performed (reasonably) well in the experimental inclusive setting.
evaluated herein. Of these, 58% were regarded as functioning above average regarding the social–emotional development and their behaviour at school and there was a positive assessment on the quality of their education. That a comfortable majority of the pupils were doing well and being educated properly seems a good score for an experimental inclusion project, but in daily practice, most attention will be focused on the 29% where development and quality of education fall short.

A number of remarks can be made regarding the interpretation of this percentage. The present study chose not to compare the experimental setting with segregated special education (see above). The evaluation of an experimental inclusive setting can easily end up in a contest between regular and special education. The fact that 29% of pupils in the experimental inclusive setting do not function as expected and/or their education is being criticized does not imply that segregated special education for these pupils would have been a better option. These pupils have not attended a special school and any speculation on what their development would have been there is without foundation. And even if such a study had been conducted, it would most likely — as many of the studies comparing regular and special settings show (Nakken & Pijl, 2002; Pijl et al., 2003) — come up with quite inconclusive results.

The second remark addresses the assessment procedures. This study asked assessors to evaluate social–emotional functioning, pupil behaviour and the quality of education of up to seven pupils on a five-point scale. Assessors in this sort of decision-making situation will tend to rank order the pupils: one pupil within the subgroup evaluated as functioning on a low level, a group of three to four pupils as average and one pupil as performing very well. The method used in this study invited assessors to create some diversity and it is possible that part of the 29% (as well as the pupils at the other end of the spectrum) can be seen as an artefact of the research design.

A last remark concerns the reference population for the assessors. As explained above, the majority of the assessors comprised peripatetic teachers with experience in supporting pupils with special needs and their teachers in regular education settings. In the Dutch context, these are in fact the only experts able to assess what can be expected of the development of pupils in inclusive settings and their education. Research (Claessens et al., 1989) has shown that the selection criteria for pupils supported with peripatetic teaching is rather high. Participation in peripatetic teaching is only for pupils for whom it is expected they will feel good and perform well in regular education. This implies that the reference population for the assessors in this study comprised relatively good performing pupils going from special to regular education. That selection mechanism does not apply for the experimental inclusive setting described herein. It is more than likely that some of the pupils with special needs in regular schools in Almere would not meet the criteria for peripatetic teaching. This makes it understandable that the assessors were worried about the social–emotional development and/or the behaviour and/or the education of some of the pupils.

These three remarks prohibit taking the percentage of pupils with a (very) disappointing development and/or quality of education simply as a shortcoming of the Almere project. Comparisons with other datasets are not available and/or would not
yield useful interpretations. The results of this study have implicitly been compared with the assessors’ knowledge and experience about the development and education of pupils with comparable impairments generally. A better-informed team of assessors is not available in the Netherlands. If one trusts their assessment, the experimental inclusion project in Almere offers good quality education to a large group of pupils and the vast majority of these are performing and developing well. The assessors are not comfortable with the situation regarding 29% of the pupils. It is obvious that the results of this study urge the teams in Almere to analyse critically the situation of those pupils and to improve ongoing assessment of pupils’ developments and the quality of education. Based on the assessors’ evaluations, the focus then should be on the social–emotional development of the pupils, as this was the most frequently mentioned reason for a negative assessment.

The analyses of qualitative data based on interviews with parents and teachers of the same pupils show they evaluated the development of their children/pupils more positively. Both parents and teachers are much more positive about the pupils’ development and education compared with the independent assessors. The large majority of parents and teachers came to a positive evaluation. That parents and teachers have more positive assessments compared with more independent assessments is not an unusual outcome (Whitney et al., 1994; Scheepstra et al., 1999; Monchy et al., 2004). However, the systematic underrating of the problems experienced by pupils is a worrying phenomenon. If parents and teachers fail to see the difficulties faced by their children/pupils, they will not be aware of the need to take action and will not do anything or act too late. This once more stresses the need to evaluate the development of pupils continuously and critically, even if all parties involved are sure that all is going well.

Notes on contributors

Sip J. Pijl is senior staff member at the Institute for Special Needs Education at the University of Groningen, the Netherlands, and also works for the European Agency for Development in Special Needs Education, Middelfart, Denmark. Dorien Hamstra is working on her PhD at the Research Institute for Educational Research (GION), University of Groningen, the Netherlands. Her thesis focuses on the evaluation of an inclusion project.

References


