DOCUMENT RÉSUME

ED 388 418 PS 023 702

AUTHOR van der Aalsvoort, Diny G. M.; Bakker, Joep T. A.
TITLE Teacher Evaluation of Academic Performance with Third Graders in Regular Education.
PUB DATE Aug 95
NOTE 16p.; Paper presented at the EARLI Conference (Nijmegen, Netherlands, August 26-31, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Elementary School Students; Foreign Countries; Grade 3; Interpersonal Relationship; Primary Education; *Student Behavior; *Student Evaluation; Student Improvement; *Teacher Behavior; *Teacher Influence; *Teacher Student Relationship
IDENTIFIERS Netherlands

ABSTRACT

This study investigated the changes in teacher evaluation of students after having participated with them in a scholastic task, and showed the relationship between evaluation and teacher behavior towards the students in regular and special education, and towards children in preschool in the Netherlands. Subjects evaluated by their teachers were 220 third graders from 10 regular education schools. Four students from each school were selected to perform a research task with their teacher which consisted of five problems from grade four mathematics books. The performance was videotaped and analyzed using the following measurements: the social support of the teacher; the competence of the child; the quality of arithmetic instruction; the regulative behavior; and the mediation quality. The results showed that there were significant differences on the evaluation ratings before and after the intervention in favor of the low-rated children. The discussion focuses on whether the changes in the evaluation of academic performance after the intervention are related to the confrontation with the actual performance or to changes in the self-fulfilling prophecy since the factor structure of the evaluation changed only in the case of the pupils rated as low. Contains 12 references. (AP)

Teacher evaluation of academic performance with third graders in regular education

Diny G.M. van der Aalsvoort
University of Leiden
Department of Special Education
Wassenaarseweg 52, Postbus 9555
2300 RB Leiden
the Netherlands
fax: 3171273619
E-mail: aalsvoor@rulfsw.LeidenUniv.nl

Joep T.A. Bakker
University of Nijmegen
Department of Special Education
Postbus 9133
6500 HE Nijmegen
the Netherlands

Paper presented with poster at EARLI Conference in Nijmegen, the Netherlands: August 26-31, 1995

Abstract

The study investigated whether changes could be brought about in teacher evaluation. Studies on teacher evaluation show the relationship between evaluation and teacher behaviour towards the pupils in regular and special education, and towards children in preschool. Evaluation not only attributes to the actual behaviour of the adult and the pupil, but in the long run also contributes to the behaviour of the child during social interaction with other teachers as well. In this study we answered the question whether teacher evaluation on children changes after having participated with them in a scholastic task. 220 third graders from 10 schools for regular education were rated by their teacher. Four pupils from each school were selected to perform a research task with their teacher: the two rated as highest and the two rated als lowest compared with their peers in the classroom. This task consisted of 5 sums from grade four mathematic books. The performance on the research task was videotaped. The videotapes were analyzed by means of five instruments: the social support of the teacher, the competence of the child, the quality of arithmetic instruction, the regulative behaviour, and the mediation quality. Factor analyses and analyses of variance revealed significant differences with regard to the evaluation ratings before and after the intervention in favour of the low rated children. The discussion centres on whether the changes in the evaluation of academic performance after the intervention are related to the confrontation with the actual performance or to changes in the self-fulfilling prophecy, since the factor structure of the evaluation changed only in the case of the pupils rated as low.

Introduction

Studies about teacher evaluation on pupils are generally executed to answer the question of whether it predicts academic performance of pupils in the (near) future. Brophy and Good (1974) and Bakker (1984) for example show that this evaluation is based on the expectation of the so called normality of pupils: a global evaluation based on the behaviour in the classroom only, indicates task behaviour, social behaviour and home environment. More recent studies of Bakker (1991), and Bakker and Ubachs (1993) also reveal that the more disciplined the behaviour of the children is, the more remediable their learning problems are, according to their teacher. These findings indicate that positive evaluation ensures an empathic attitude towards the child in the classroom. Research as to whether teacher assessment is connected to the child's actual performance has recently been undertaken. Van der Aalsvoort (1993) studied social interaction between preschoolers and preschool teachers in child care. Her study shows that more the preschool teacher is concerned about a specific child, the less helping behaviour and positive emotional support he/she shows towards this child. Smits (1993) found likewise that evaluation was influenced by the performance of young children. Children in at-risk groups receive less positive emotional support and are more often reprimanded during social interactions with the teacher than their peers. As the feeling of cognitive competency is also lower with the children in at-risk groups, the teacher evaluation thus becomes reality in the long run. The lowly evaluated children do indeed perform poorly as the teacher already expected.

Brophy and Good (1974), e.g., show that instruction quality is superior with high evaluated pupils compared with their peers. Their study does not include information on other forms of social interaction, such as verbal and non verbal emotional support of the teacher, or characteristics of the child while performing, as in the studies of Smits (1993) and Van der Aalsvoort (1993).

If teacher evaluation acts as a self-fulfilling prophecy, it becomes important to investiga-

:e whether negative consequences of poor evaluation can be countered. It may be argued that specific processes during task performance are responsible for the occurance of self-fulfilling prophecy. One explanation is offered by Wertsch and Sammarco (1985). In their view, teacher behaviour depends on the definition of the task goal in the specific social context of the classroom, where the teacher is responsible for the child's achievements. Better performing children understand this definition faster, so they need less regulation on the part of the teacher to reach the task goal, than their poor performing peers. This behaviour is anticipated by the teacher, and therefore he/she tends to control them less by rules and restrictions, in other words, regulations. This in turn elicits the attunement of the better performing children towards the task and enhances the stability of teacher evaluation. According to Vye, Burns, Delclos and Bransford (1987), teachers evaluate children more positively after having observed the child's performance on videotapes with other adults. Their findings offer a second explanation for teacher evaluation: the child's cognitive ability is reconsidered without having been involved with them in an academic task.

It needs further investigation whether the process described before can be countered by exposing the teacher to pupils who perform better than expected, or to children who profit more from help than he/she had anticipated. The findings stated before indicate that it is important to study both teacher and child behaviour in order to analyze how social interaction and evaluation are connected during task performance.

A study was executed to answer the question whether the evaluation of academic performance was related to the social interaction in the task situation. Based on the findings of Brophy and Good (1974), and Van der Aalsvoort (1994) it was hypothesized that lowly evaluated children receive less instruction and support and are more often regulated than their peers. Supported by the findings of Vye, et al. (1987), and Bakker and Ubachs (1993), it was also hypothesized that teachers would assess the task behaviour of the formerly lowly evaluated

children more favourably.

Method

Sample

Ten third grade teachers from ten different schools for regular education participated. They evaluated a total of 255 children, 139 boys (mean age 98 months) and 116 girls (mean age 104 months). Four pupils from each school took part in the intervention: the two rated as highest and the two rated als lowest compared with their peers in the classroom.

Procedure and instrumentation

The study was a pretest-intervention-posttest design. Teachers from ten schools were asked to join in the study. They rated all pupils by means of a questionnaire consisting of 25 items, using bipolar, 7-point rating scales (Bakker, 1984). The items were grouped by orthogonal factor analysis into three factors: task behaviour, social behaviour, and a factor of items related to home environment. The teacher was not aware of the selection criterion on the four pupils.

The intervention session took place in the classroom. During the task the teacher was seated with each pupil individually, and assisted the child with the completion of 5 sums. The order of the sums was standardized: the first and second sum were at the actual performance level of the pupil, while the next three sums seemed to be too difficult. The teacher was advised to help the child, but was free in formulating that help, and in offering arithmetic material during the task. The session with each child took about 10 minutes, or less where the sums were solved more quickly. The intervention was videotaped. The other children in the classroom carried out tasks on their own during the research session. In the week after the intervention, the teacher evaluated all the pupils again, including those in the intervention group. The question-

naire in the pretest and post test, though formulated differently, consisted of the same items.

The videotapes were analyzed by means of five current instruments, after substantial interrater reliability had been obtained (Cohen, 1960).

The first instrument, designed by Erickson, Sroufe and Egeland (1985), consisted of ratings on a seven-point scale, designed to evaluate the social support of the teacher five subscales: supportive presence, respect for the child's autonomy, structure and limit setting, hostility, and quality of instruction (rating=1: no to little evidence of support, rating=7: strong evidence of support). The second instrument was also designed by Erickson, Sroufe and Egeland (1985), and consisted of ratings on a seven-point scale, designed to evaluate the competence of the pupil with four subscales: avoidance of the teacher, reliance on the teacher, perseverance, and compliance ((rating=1: no to little evidence of competence, rating=7: strong evidence of competence). The third instrument, designed by Wertsch and Sammarco (1985), covering the use of rules and restrictions during task performance, was based on the amount of direct and indirect regulative behaviour on the part of the teacher in the task, while the child was reading, writing and solving the sums. The frequencies of the observed regulations were added. The fourth instrument, based on the theoretical model devised by Van Parreren and Carpay (1972), was designed to establish the description of instructional behaviour of the teacher. The activities to be registered were orientation, instruction, and product and process feedback on the part of the teacher. The frequencies of the instructional behaviour were added. The fifth instrument on mediation quality and designed by Lidz (1991), consists of 12 subscales: intentionality, meaning, transcendence, joint regard, sharing of experience, task regulation, praise, challenge in the zone of proximal development, psychological differentiation, contingent responsivity, affective involvement and change. Each subscale is rated from 0 (not apparent) to three (optimal mediation).

The first hypothesis was tested by t-testing the mean scores on the instruments to measure the social interaction between the highly and lowly evaluated children. The second one

was tested by orthogonal factor analyses (varimax criterion) of the evaluation ratings in the pretest and post test.

Results

The first question as to whether the evaluation of academic performance was related to the social interaction in the task, was answered by t-testing the mean scores on the instruments designed to measure the social interaction. It was hypothesized that lowly evaluated children received less instruction and support, and that they were more often regulated than their peers.

---

table 1 about here

As table 1 shows, the mean differences in teacher behaviour between the lowly and highly evaluated group were small. T-testing revealed no significant differences between the groups when the social support, the regulative behaviour, the instruction quality, and the mediation quality were compared. Likewise no differences were found when the competence of pupils during the task performance was compared. Our hypothesis that evaluation and social interaction are correlated, was not confirmed.

The second question, as to whether evaluation of academic performance was related to the intervention, was examined by comparing the factors in the dataset of the pretest with those in the post test. A multivariate analysis was carried out in order to describe the two datasets. According to the varimax criterion (Bakker, 1984) three factors emerged after an orthogonal factor analysis with rotation in the pretest and post test scores: task behaviour, social behaviour, and home environment. Table 2 shows the loadings of each factor on the pretest. The table shows that 71 % of the variance of the evaluation in the pretest is explained by the three

factors: task behaviour accounts for 46 % of the variance (65 % of the extrapolated variance).

Table 2 about here

Table 2 also reveals significant differences between item scores when the pretest and the post test scores of all subjects are compared.

The means of the highly evaluated group compared with the lowly rated group differed significantly between the groups before the intervention: both the task behaviour ($t=17.12$, $p<.00$), the social behaviour ($t=12.54$, $p<.00$), and the home environment ($t=9.79$, $p<.00$). These findings point in the direction of a more positive evaluation of the highly rated pupils.

In the post test again the highly evaluated pupils were rated significantly better than their lowly evaluated peers. Both the means on the disciplined behaviour ($t=9.37$, $p<.00$), the home environment/task behaviour ($t=8.15$, $p<.00$), as well as the mixed factor ($t=9.08$, $p<.00$) differ significantly between the groups. In the case of the lowly evaluated group, the scores on items related to task behaviour ($t=2.52$, $p<.01$) were significantly lower on the post test, compared with the pretest.

Discussion

The results on the first question examined during this study, whether there is a relationship between teacher evaluation of academic performance and social interaction, could not be confirmed. Although teachers rated their pupils differently, as the results reveal, they did not behave differently towards the children. This outcome is in contrast with the findings of Brophy and Good (1974). Further analyses will be needed to establish whether differences in social support, instruction quality, regulative behaviour, and mediation quality are related to differences in

evaluation.

One explanation is that evaluation and subsequent social interaction is sex specific. Findings of Van der Aalsvoort (1994) suggest that even at two and three years of age there is a difference between the support which the adult gives to boys than that which they give to girls. In fact Wagenaar and Scholte (1991) found that boys receive more regulative remarks from their teacher than girls. Another explanation is that the change in attitude is more marked in the case of boys. To test this hypothesis, an analysis on the level of evaluation statements is needed.

Our findings with respect to the second question reveal that teachers change their opinion of pupils in the classroom after being confronted with the actual behaviour of these children in a shared task. Items that referred to the actual behaviour during the research task, such as 'thinks impulsively' and 'seldom sits still' were especially sensitive to a change of evaluation. This result is noteworthy since studies on teacher evaluation thus far show that teachers evaluate globally, suggesting that they do not change their opinion easily. Our findings however confirm that teachers, when asked to evaluate their expectation of a child, actually evaluate the behaviour of that child. Teacher attitude towards pupils therefore seems to be based on task behaviour in the classroom.

Another remarkable result is the fact that the ratings of the teacher improved only with those pupils who were evaluated lowly before the intervention. According to our findings lowly and highly pupils looked more 'alike' to the teacher after having shared a task situation with them. As it was found that for both groups, the highest loaded factor in the post test was 'disciplined behaviour', the conclusion was, that the change in evaluation must be attributed to the intervention, since the evaluation of the total group did not change.

It remains unclear whether the change in evaluation is due to the behaviour of the highly evaluated pupils or to the behaviour of the lowly rated ones. It could be argued that the teacher adjusts his or her opinion after having been confronted with four task oriented children in a row.

The high loading on the factor 'disciplined behaviour' suggests that all subjects from the intervention group made the teacher more optimistic about their remediability. It is more likely however, that the teachers regarded task behaviour of the pupil as belonging to their effort. Other studies on teacher evaluation confirm these results (Brophy & Good, 1974; Bakker, 1991; Bakker & Ubachs, 1993). In the case of children rated as high, the evaluation of teachers more likely attribute this to the child's performance, while in the case of pupils evaluated as lower, the teacher attributes this to factors outside the classroom.

In this study the confrontation with the actual task behaviour of lowly evaluated pupils altered the teacher's opinion of them. Further research will be needed to establish whether this relationship between teacher evaluation and teacher behaviour in the classroom is indeed a causal one. This point is reinforced by the fact that in this investigation, the teachers had altered their opinion about pupils previously rated as low, after sharing a task performance with them.

References

Bakker, J.T.A. (1984). De leerkracht en de slechte leerling [The teacher and the poor performing pupil]. Meppel, Krips Repro: proefschrift.

Bakker, J.T.A. (1991). Teacher evaluation of children with learning problems in regular and special education. Tijdschrift voor Orthopedagogiek, 30, 465-475.

Bakker, J.T.A., & Ubachs, C.B.C.M. (1993). Teacher perception of chances of remediation. Tijdschrift voor Orthopedagogiek, 32, 19-29.

Brophy, J.E., & Good, T.L. (1974). Teacher-Student Relationships. New York: Holt, Rinehart & Winston.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational Measurement, 20, 37-46.

Erickson, M.F., Sroufe, L.A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high risk sample. In J. Bretherton & E. Waters (Eds), Growing points of attachment theory and research. Monographs of the Society for Research in Child Development, 50, (12 serial, no. 209).

Smits, S.C.M. (1993). Pedagogical antecedents of on-task behavior and achievement in school performance. Utrecht, ISOR: proefschrift.

Van der Aalsvoort, G.M. (1993). The quality of social interactions during performance of learning potential tasks with preschoolers and their caretakers. Tijdschrift voor Ontwikkelingspsychologie, 20, 2, 247-262.

Van Parreren, C.F., & Carpay, J.A.M. (1972). Sovjetpsychologen aan het woord. [On Soviet Psychology]. Groningen: Wolters-Noordhoff.

Vye, N.V., Burns, M.S., Delclos, V.R., & Bransford, J.D. (1987). A comprehensive approach to assessing intellectually handicapped children. In C.S. Lidz (Ed.), Dynamic Assessment. An interactional approach to evaluating learning potential. New York: Guilford Press.

Wagenaar, E., & Scholte, E.M. (1991). Teacher-pupil interactions, pupil-characteristics and learning capacity of young Turkish and Moroccan children. Tijdschrift voor Orthopedagogiek, 30, 228-239.

Wertsch, J.V., & Sammarco, J.G. (1985). Social precursors to individual cognitive functioning: the problem of units of analysis. In A. Hinde, A. Perret, S. Clermont, & J. Stevens-Hinde (Eds.), Social relationships and cognitive development. Oxford: Clarendon Press.

Table 1: Means, standard deviations and results on t-testing the data of the social interaction during the intervention of the lowly and highly evaluated pupils.

	low evaluation		high evaluation	
	M	SD	M	SD
Instruction quality	6.25	2.24	6.80	1.94
Regulative behaviour	5.40	1.96	4.65	2.16
Social support	27.60	5.67	30.70	5.15
Competence	20.55	4.22	22.35	3.01
Mediation Quality	17.35	4.84	17.45	3.61

Table 2: Factor patterns on evaluation of all subjects (N=255) in the pretest and in the posttest (underlined numbers)

Factor 1: Task behaviour

Factor 2: Social behaviour

Factor 3: Home environment

Statement on the child	Factor 1	Factor 2	Factor 3

2.	Dreams often	.84	**.81**	.16	**.19**	.14	**.25**
4.	Seldom keeps promises	.63	**.72**	.42	**.35**	.28	**.13**
7.	Is easily distracted	.83	**.73**	.19	**.36**	.19	**.20**
8*.	Needs frequent warning	.68	**.55**	.46	**.60**	.03	**.13**
12.	Delivers superficial jobs	.82	**.84**	.26	**.18**	.18	**.18**
15.	Thinks impulsively	.70	**.75**	.41	**.31**	.27	**.11**
16.	Deviates from instruction	.62	**.78**	.45	**.28**	.06	**.16**
18.	Seldom sits still	.72	**.65**	.30	**.38**	.18	**.18**
1.	Gets angry easily	.21	**.22**	.85	**.81**	.02	**.10**
5.	Laughs at mistakes from others	.27	**.32**	.84	**.79**	.12	**.14**
10.	Hardly gives in	.29	**.31**	.76	**.78**	.10	**.15**
11*.	Is meddlesome	.48	**.59**	.48	**.54**	.08	**.15**
14.	Teases other children	.24	**.26**	.84	**.83**	.09	**.13**
19.	Wants to have his way	.50	**.51**	.67	**.68**	.20	**.17**
3.	The parents hardly discern toys	.14	**.08**	.09	**.08**	.84	**.87**
6.	The family is instable	.34	**.35**	.02	**.13**	.64	**.58**
9.	They rarely visit museums	.13	**.20**	.00	**.07**	.88	**.75**
13.	They are poorly educated	.22	**.21**	.05	**.05**	.84	**.83**
17.	The child looks too much tv	.04	**.11**	.18	**.24**	.80	**.79**
20.	The parents ignore adequate language behaviour	.07	**.05**	.10	**.12**	.87	**.83**

* Item 8: belonged to the second factor in the post test
* Item 11: belonged to the first factor in de post test