

Take it from the Bright and Give it to the Poor? Some Remarks on Within-Classroom Homogeneity in Reading Achievement

Rainer H. Lehmann

*Humboldt-University Berlin
Department of Education*

Abstract

It is argued in this paper that within-classroom homogeneity of achievement is an important criterion of successful instruction. Data from the German component of the International Reading Literacy Study (IEA) are investigated to demonstrate this point. 251 randomly sampled intact classrooms from grade 3 and 297 classrooms from grade 8 (total $N > 11.000$) are considered in the analyses. It is shown that moderate degrees of within class heterogeneity are often associated with high average performance in primary school reading achievement. Findings from the lower secondary level are mixed. As it seems, the intention to remove disadvantages for the weaker students does not, in itself, guarantee higher achievement, nor is it necessarily linked to the motivational advantages which mixed-achievement groups can provide.

1 Introduction

When educational achievement is investigated, it usually goes without saying that *mean* achievement is considered as the primary quality index for classrooms (teachers), schools, school types, or educational systems. There are good reasons, however, to investigate also the degree to which students of the same unit or system reach similar levels of achievement. The concept of "mastery learning" (Bloom 1968; Block 1971), for instance, was introduced on the basis of the assumption that it is possible to attain high average performance levels by way of instructional techniques which facilitate growth especially among the weaker students. Expressed in more abstract terms, this approach posits that the quality of an educational (sub-)system would also manifest itself in a reduction of *variance* within the group of learners.

Obviously, though, it cannot be taken for granted that this strategy be void of 'opportunity costs', paid for by brighter students who achieve below their potential. On the other hand, it might also be the case that fast learners actually profit more from "mastery learning" than do the slower ones (Weinert 1982).

The former two Germanies render an interesting case for this debate. Post-World War II history has resulted in the emergence of clearly diverging educational policies, the differences of which can be traced exactly to this question of how best to deal with heterogeneous 'starting conditions' among students. The 'conservative' West German states ("*Länder*") have retained the traditional, socially selective tripartite system of secondary schools and restricted access to the more privileged tracks, believing that the overall yield could be maximized by arranging relatively homogeneous groups of learners. As opposed to this, 'liberal' reform initiatives in West Germany have sought to equalize educational opportunities on the basis of a mixed system which involves increased admittance to the academic and intermediate tracks in the tripartite framework, as well as "comprehensive schools" operating with some degree of external differentiation in core subjects. East Germany, finally, has opted for a true comprehensive system up to grade 10, with a highly selective transition to the 'academic' grades 11 and 12 (for details, see Führ 1979; Lehmann 1994). Given that there is a long common history and, in spite of the ideological controversies which lasted for almost half a century, also much commonality in the general cultural background, it is of considerable interest to study the effects of the divergent policies not only on averages of achievement, but also on its homogeneity. At the same time it will be instructive to see how these two criteria are interrelated.

In a recent comparative study of educational achievement and its determinants in East and West Germany (North Rhine-Westphalia only), it has been observed that, in terms of the overall achievement distribution in grade 7, the East German students display slightly better average performance, with clear advantages at the lower end and without significant draw-backs among the top students (Max-Planck-Institut für Bildungsforschung & Institut für die Pädagogik der Naturwissenschaften 1994, pp. 25ff.). Since the East German curriculum had allowed more instructional time in all subjects tested (mother-tongue, mathematics, physics, biology), it was suggested that the better performance of weak East German students may, indeed, have been facilitated by compensatory teaching strategies and additional instructional time (op. cit., p. 28). This finding coincides generally with results from the IEA Reading Literacy Study (Elley 1992), where the former East Germany and the 'conservative' West German

Länder were found to be superior in terms of average achievement to the ones with a less selective approach, (Lehmann, Peek, Pieper & von Stritzky 1995, pp. 143ff.). Remarkably, the policy of 'compensatory education' in the 'liberal' traditionally Social-Democrat states had not resulted in a significant reduction of the overall variance.

It can be argued that the primary mechanisms behind these observations are linked only indirectly to educational policy at the state level. Selectivity of the system as a whole (or the reduction thereof) translates into general educational aspirations and subsequently into concrete classroom processes. At both levels, much will depend on the frame of reference within which the individual learners perceive and define their own positions: many variables known to be closely related to successful learning - patterns of self-concepts and causal attribution, expectations of success and choices of values, and above all learning motivation - are likely to be influenced heavily by the success of others in the same learning group. At the very least, effects of this kind will have to be considered when the significance of classrooms is emphasized (as, for instance, in Ditton & Krecker in press).

These considerations underline, from a more theoretical perspective, that the role of within-classroom homogeneity should be analyzed more extensively than has been the case so far. While, as in the case of the IEA Reading Literacy Study, research on the effectiveness of schools (Postlethwaite & Ross 1992) and the influence of school and teacher variables (Lundberg & Linnakylä 1993) was certainly necessary and conducive to a better understanding of the factors behind the development of high levels of reading comprehension, these studies are still conceptually and analytically tied to the notion of mean achievement. Thus, it may be helpful to supplement these analyses by an exploratory investigation of possible moderating effects associated with the achievement variation within the learner group.

Because of its rather unique political context, only briefly outlined above, the German IEA Reading Literacy Study appears to provide a particularly suitable data base for this purpose. Therefore, the following will be restricted to some selected findings from these data. Also, reading comprehension as a dimension for achievement has some distinct advantages. As measured in this study, it represents a fairly unidimensional trait which is certainly influenced by instruction, but which is also likely to be a function of underlying general abilities (Lehmann et al. 1995, p. 42). In addition, it would be quite inappropriate to consider reading comprehension as the unique product of a given classroom (teacher), but it can serve as a useful measure for the general academic orientation of students and classrooms.

If the psychological effects suggested above do, in fact, occur, it is not plausible to expect a tendentially linear relationship between within-classroom homogeneity and mean achievement. There are substantial reasons to hypothesize a curvilinear model: The theory of achievement motivation would imply that an intermediate level of challenge, as is likely in a moderately homogenous classroom, will be associated with high motivation and hence achievement (Heckhausen 1967). Also, the (in-)ability of teachers to manage heterogeneous classes will put noticeable constraints on achievement.

2 Data Structure and Methodology

The general design of the IEA Reading Literacy Study has been described elsewhere (Elley 1994). Suffice it to say here that two populations were assessed: predominantly 9-years-olds (in Germany: grade 3) and predominantly 14-year-olds (grade 8). Internationally the reading tests required roughly 90 minutes time, which, in the German case, was increased by another 45 minutes for additional items. This implies that the German national scales, if computed on the basis of the international plus the national item pool, differ from the international scales. In order to avoid confusion, these national scales are referred to in terms of logit scores (derived from the one-parameter Rasch model: Wright, Linacre & Schulz 1990), which are not directly comparable to (though highly correlated with) the international scales in the official IEA publications.

The German samples consisted of randomly drawn intact classrooms, with a stratification scheme taking account of the number of students in the target grade by state and (where applicable) by school type. The exact figures are given in Table 1. If not indicated otherwise, no weights were used for the present analyses.

In order to study the relationships between within classroom homogeneity and reading achievement, student scores were aggregated to the classroom level. Apart from the respective averages, within-class standard deviations and variances (uncorrected for estimation bias) were computed.

Table 1: Sample Structure of the German Component of the IEA Reading Literacy Study.*

Subsample	No. of classes	No. of students
Grade 3, East	101	1,882
Grade 3, West	150	2,958
Grade 8, East	100	1,873
Grade 8, West; of these	197	4,348
- lower track (" <i>Hauptschule</i> ")	67	1,327
- intermediate track (" <i>Realschule</i> ")	54	1,227
- academic track (" <i>Gymnasium</i> ")	58	1,379
- comprehensive (" <i>Gesamtschule</i> ")	18	415

* The psychometric properties of the reading achievement measures used were satisfactory; they ranged from $\alpha = .89$ (grade 8, East) to $\alpha = .93$ (grade 3, East and West).

The hypothesized curvilinear relationship between average achievement and within-classroom homogeneity can then be captured by the following multiple regression model:

$$x_j = b_0 + b_1 * s_j + b_2 * s_j^2 + error$$

with x_j = mean reading achievement in classroom j
 s_j = standard deviation of reading achievement in classroom j
 s_j^2 = variance of reading achievement in classroom j
 $b_{0,1,2}$ = regression coefficients

The underlying hypothesis leads to the expectations of $b_1 > 0$ and $b_2 < 0$. As in many regression models of this type, there will be some unavoidable problems of multicollinearity; these appear acceptable, as long as the curvilinear model is indeed superior statistically to a simpler one using either the standard deviation or the variance as the single predictor.

3 Results

When searching for classroom-level covariates of reading achievement, it appears most promising to start with the lower grades in the educational system, because here context conditions are likely to have been relatively

constant up to the point of measurement. Accordingly, the findings for the two samples of 9-year-olds are presented first (see Tab. 2).

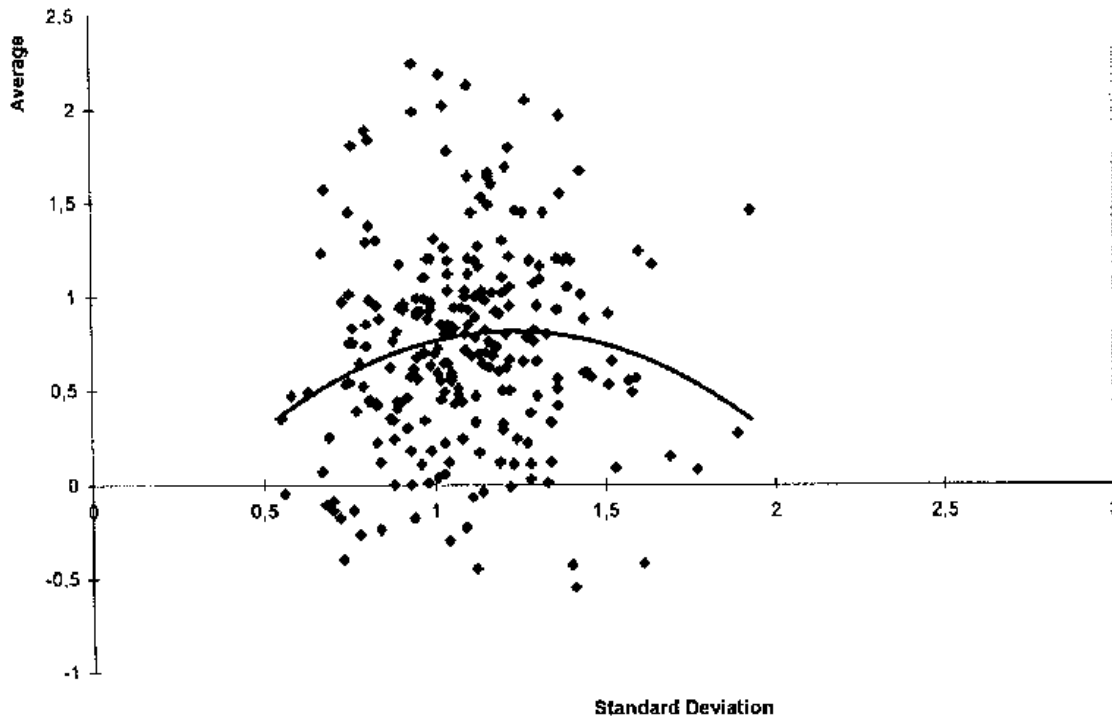
Table 2: Multiple Regression of Average Reading Achievement in Grade 3 on Within-classroom Homogeneity (Standard Deviation and Variance); East and West Germany.

Sample	<i>N</i>	Total Mean	Total <i>SD</i>	b_0	b_1 (linear)	b_2 (quadratic)	R^2	p (<i>F</i>)
East Germany	101	.70	1.25	-.77	2.65	-1.14	.033	.19
West Germany	150	.78	1.26	-.58	2.22	-.86	.033	.08

As can be seen, the West German students are slightly superior to their East German peers in terms of overall performance, but this difference is statistically insignificant if corrected for "design effects" (Kish 1965). There is virtually no difference with regard to the overall standard deviation, and the same is true for the pooled-within-class standard deviation. So, in spite of the stronger adherence of East German teachers to the methods of "direct instruction" (Lehmann et al. 1995, pp. 69ff.), similarities rather than differences between the two parts of the country stand out. The same is true for the two regression equations. Although neither of them quite reaches statistical significance, they do confirm the theoretical expectations, and for the two files combined, all parameters (except the additive constant) as well as the overall *F* are, indeed, significant ($p(F) = .02$).

If average reading achievement is plotted against the within-class standard deviation, a remarkable pattern emerges (see Fig. 1): The upper boundary of the plotted points is shaped like an inverted parabola, whereas the lower boundary follows more or less a straight line. The density of the plotted points also depicts a curvilinear trend. This suggests that a moderate amount of within-classroom homogeneity is a necessary, though not sufficient condition for successful reading instruction from which, on the average, the children profit the most.

Figure 1: Average Reading Achievement Plotted Against Within-classroom Homogeneity (Standard Deviation) in 251 Classrooms, Grade 3, from East and West Germany.



At the lower secondary level, the situation is bound to be much more complex. First of all, at this stage, reading achievement cannot sensibly be considered as the product of a single classroom setting. Especially under the condition of the West German system of tracking students at the end of grade 4, this assumption is clearly inappropriate. Secondly, the respective selective processes already entail homogenizing effects which will have to be considered. Finally, it is conceivable that classroom homogeneity is related to mean achievement in different ways at different levels; note that school type and reading achievement are highly inter-related ($\eta = .64$)! Table 3 presents the respective results; because of their small number, comprehensive schools in West Germany are omitted.

Table 3: Multiple Regression of Average Reading Achievement in Grade 8 on Within-classroom Homogeneity (Standard Deviation and Variance); East Germany and West Germany by Policy and School Type.

Sample	N	Total Mean	Total SD	b_0	b_1 (linear)	b_2 (quadratic)	R^2	p (F)
East Germany	100	1.49	1.01	1.40	.24	-.13	.00	.83
West Germany ('liberal' states)	108	1.31*	1.08*	-	-	-	-	-
- lower track	33	.47	.85	1.28	-1.54	.54	.17	.06
- intermediate track	30	1.26	.71	2.94	-5.18	3.88	.04	.59
- Academic track	35	2.20	.84	2.35	.01	-.28	.06	.36
West Germany ('conservative' states)	89	1.47*	1.11*	-	-	-	-	-
- lower track	34	.62	.86	2.15	-4.41	3.04	.04	.57
- intermediate track	24	1.70	.75	1.19	1.73	-1.44	.00	.96
- Academic track	23	2.39	.87	-.96	7.88	-4.29	.33	.02

* Weighted by state and school type.

As compared with the findings from the younger group, these data may, indeed, seem disillusioning. They confirm, of course, the homogenizing effects of the tracking system on schools and classrooms. They also include the intriguing result that the West German states with a highly selective policy obtain an average performance level which is about as high as the respective value for the comprehensive East German school system; it is, in fact, slightly higher, if only students are considered whose mother-tongue is German. The West German states which emphasize equality of educational opportunity, on the other hand, fall behind, and this cannot be accounted for by their share of immigrant students. To this extent, the findings from the MPI/IPN study are confirmed at the macro level. The fact, however, that the former "Secondary Polytechnic Schools" in East Germany combine their commendable performance levels with less variation in the system as a whole, is due to the absence of immigrant students. In turn, high selectivity as present in the 'conservative' West German "Länder" is perceptible in a slightly increased overall standard deviation, if only non-immigrant students are considered.

The question with which this paper is primarily concerned necessitates a closer look at the regression coefficients for the various subsamples. The

small sizes of the latter as well as the above-mentioned theoretical constraints call for not overemphasizing the criterion of statistical significance here. It seems more important to note that there is, indeed, a pattern in the matrix of regression coefficients.

Ironically, in the West German "*Länder*" with a less selective approach, a tendency for only the more homogeneous classrooms to be associated with higher mean performance prevails throughout the three different tracks. In fact, the simple regression model using only the standard deviation as a predictor always has a better model fit. The probabilities for type I errors ($p(F)$) are: lower track .02; intermediate track .43; academic track .16. In other words: the overall yield is high only where the homogenizing effects are strong. Taken together, these observations contradict two tenets of the 'liberal' approach: Obviously, expanded access to the more privileged tracks in the system has not significantly reduced the overall variance, let alone increased the overall performance level. Instead, increased within-classroom heterogeneity where it has occurred, seems to be associated with generally lower achievement.

By contrast, no such tendency is present either in the former East German schools or in those of the more selective West German "*Länder*". For the most part, within-class homogeneity is virtually unrelated to mean achievement. The only exception is given by the academic track ("*Gymnasium*") in 'conservative' West German states where the parabolic distribution, which was characteristic of the younger population, persists. Apparently, "academic" classrooms can operate very successfully at moderate degrees of heterogeneity in these states for which a policy of setting clear, centralized standards is characteristic.

4 Discussion

Some evidence has been found which confirms the hypothesis that high within-classroom mean achievement is associated with moderate levels of heterogeneity. From a strictly statistical standpoint, the most pronounced homogeneity would be expected to be found in classrooms with extremely high or low average performance, if (as in this case) no bottom or ceiling effects are present. This clearly does not apply to the distributions encountered. Quite to the contrary, the best performing 3rd-grade classrooms (and also the highest achieving classrooms in the traditional "*Gymnasium*" in 'conservative' West German states) are characterized by an intermediate

range of achievement (and ability?). At the same time, relatively low average performance can be encountered at all levels of within-classroom homogeneity. In that sense, a moderate homogeneity level was suggested to function as a necessary, but not sufficient condition for successful instruction.

Several factors may contribute to this finding. (Quite understandably), teacher inability to cope with extreme achievement variation in the classroom may partly account for the absence of high-performing, yet heterogeneous classes, but it cannot explain why few or no uniformly high-achieving classrooms appear. So, the theory based on frames of reference renders, perhaps, the more plausible explanation. If student performance is significantly influenced by peer achievement, both the stronger and the weaker students might profit from comparisons with their classroom neighbors: the brighter students, because their self-concept is strengthened by their relative position in the classroom, and the weaker ones, because the (current) average performance level could serve as an attainable goal. These effects, if they exist, could be reinforced by teaching strategies which are geared to the average, but close enough to the current state of the slower learners and still stimulating enough for the faster ones. If this interpretation is correct, the present results provide evidence against both extremely homogeneous and extremely heterogeneous learner groups. The former would deprive many students of stimulating opportunities arising from comparisons with their peers; the latter would render such comparisons demotivating and, very likely, lead to a point where the teacher can no longer handle the variation encountered in the classroom.

There is, indeed, some external evidence to strengthen this point. Horstkemper (1987) has found, in a study of "tracking" vs. "setting" in comprehensive schools, that the self-concept of weaker students is more positive in the moderately homogeneous "tracked" classes. Thus, the relationships between psychological variables of this type and achievement may also help to explain why, at the lower secondary level, a more 'liberal' educational policy seems to go along with generally lower achievement levels, as well as a remarkably different joint distribution of within-classroom homogeneity and achievement.

As will be recalled, these subsamples displayed the common pattern of a negative relationship between the within-class standard deviations and the means of achievement: tendentially, high levels of homogeneity coincide with high performance and the more heterogeneous classrooms show an unsatisfactory general level. It is far from being clear why, under the conditions of a more 'liberal' educational policy, this should be so. The present data do not provide easy answers. One might speculate that clarity of

learning objectives (which is a necessary prerequisite for the assumed motivational mechanisms to function) is fostered more under the auspices of a traditional system with more elements of central control (e.g., centrally administered school graduation examinations; more powerful inspectorate). In the former case - which is not to be verified here - the factor of teacher (in-)ability to manage heterogeneous classrooms may become the dominant force. Here, too few teachers appear to be able to foster growth among the poorer students without taking it from the bright.

From the standpoint of policy-making, it is at least as interesting to consider the variance components associated with school types and the classrooms themselves. A policy geared towards equality of educational opportunity by opening access to the more privileged tracks in the system would normally be expected to increase these variance components. However, this has clearly not taken place in the 'liberal' West German states: in fact, both components seem to be slightly reduced, as compared with the 'conservative' ones. (Incidentally, this rules out the possibility that the difference in regression patterns is just a consequence of a shift towards more heterogeneous classrooms.) However unexpected this phenomenon, one might still hope that the increased homogeneity within school types and classrooms would provide a basis for more successful instruction. Unfortunately, the data do not confirm this expectation either; in the 'liberal' West German states, performance levels are lower in all school types as well as in the system(s) as a whole.

It is a widely accepted notion that achievement deficits in the lower tracks of the (West) German secondary school system are due to "creaming effects" (Tillmann 1983, 1988). As the relatively brighter move, in a higher proportion, to the more privileged tracks, they leave behind groups which display lower school type averages. The current data suggest, however, that this interpretation captures at best only part of truth. The overall average should, in this perspective, remain unchanged or even rise due to additional learning opportunities, which is clearly disproved by the evidence. If, however, the within-classroom variance drops below the optimum in the lower tracks and is increased beyond it in the academic track, it may well happen that the system as a whole suffers.

It remains to be mentioned that some of the subsamples (the former Secondary Polytechnics in East Germany; the lower and intermediate tracks in the 'conservative' parts of West Germany) show no relationship between within-classroom homogeneity and mean achievement. The most noteworthy features here are that the East German distribution shows the greatest variation among relatively homogeneous classes and almost average levels at

the heterogeneous end, whereas in the other two subsamples high performance levels also occur in heterogeneous classes. Both instances suggest that some successful instruction must have taken place in classrooms with students representing rather widely varying ability levels.

5 Conclusion

The analyses presented here are not based on particularly strong evidence. There are weak statistical points such as a lack of statistical significance in some cases; in others, the absence of correlations has been the starting point for some speculation. Nevertheless it is hoped that this first exploratory scan for information hidden in the variance components of the German IEA Reading Literacy Study, especially the ones indicating within-classroom homogeneity, has demonstrated the usefulness of such investigation.

One of the most urgent questions to be resolved in subsequent studies is that of a true control of input mean and variance as opposed to the outcome statistics. Here, reading achievement had to serve in both functions, but in the longer run it is, of course, necessary to use two distinctly independent measures.

The present findings do suggest that such an endeavour may well be worthwhile. It is conceivable that multi-level analysis techniques be extended in a way so as to use variances as predictors, if not predictands. Substantially, the evidence encountered indicates that, if applied with prudence, instruction in mixed-ability groups is particularly well compatible with high mean achievement. Although this approach may include the risk of taking potential cognitive growth from the bright when trying to give it to the poorer learners, this is not a necessary outcome. What has also become clear, however, is that the mere intention to equalize educational opportunity can, ironically, be paired with an implicit bonus for homogeneously high achieving learner groups and thus counteract its own principles.

The data used for this analysis were collected in a research project funded by the German Federal Ministry of Education and Science: the German Reading Literacy Study which, in turn, forms part of the International Reading Literacy Study conducted by the *International Association for the Educational Achievement* which was coordinated by T. Neville Postlethwaite. Funding and, even more so, the intellectual leadership provided by the International Coordinator, are gratefully acknowledged.

Bibliography

- Block, J.H. (Ed.). (1971). *Mastery Learning: Theorie and Practice*. New York: Holt, Rinehart & Winston.
- Bloom, B.S. (1968). Learning for Mastery - Instruction and Curriculum. *Evaluation Comment*, 1 (2), 1-12.
- Ditton, H. & Kreckler L. (in press). Qualität von Schule und Unterricht - Empirische Befunde zu Fragestellungen und Aufgaben der Forschung [To be published in: *Zeitschrift für Pädagogik*].
- Elley, W.B. (1992). *How in the World do Students Read? - IEA Study of Reading Literacy*. Hamburg: International Association for the Evaluation of Educational Achievement.
- Elley, W. (Ed.). (1994). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford: Pergamon.
- Führ, C. (1979). *Education and Teaching in the Federal Republic of Germany*. Bonn: Inter Nationes.
- Heckhausen, H. (1967). *The Anatomy of Achievement Motivation*. New York: Academic Press.
- Horstkemper, M. (1987). *Schule, Geschlecht und Selbstvertrauen - eine Längsschnittstudie über Mädchensozialisation in der Schule*. Weinheim: Juventa.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Lehmann, R.H., Peek, R., Pieper, I. & von Stritzky, R. (1995). *Leseverständnis und Lesegewohnheiten deutscher Schülerinnen und Schüler*. Weinheim: Beltz.
- Lehmann, R.H. (1994). Art. "Germany: System of Education". In T. Husén & T.N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (2nd ed.), vol. 5, pp. 2470-2480. Oxford: Pergamon.
- Lundberg, I. & Linnakylä, P. (1993). *Teaching Reading Around the World - IEA Study of Reading Literacy*. Hamburg: International Association for the Evaluation of Educational Achievement.
- Max-Planck-Institut für Bildungsforschung & Institut für Pädagogik der Naturwissenschaften (Eds.). (1994). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter. Zwischenbericht für die Schulen*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Postlethwaite, T.N. & Ross, K.N. (1992). *Effective Schools in Reading: Implications for Educational Planners*. Hamburg: International Association for the Evaluation of Educational Achievement.
- Tillmann, K.-J. (1983). Schülerlaufbahnen, Abschlüsse, Chancengleichheit. Anmerkungen zu einem merkwürdigen "Konsenskapitel" im BLK-Gesamtschulbericht. *Die Deutsche Schule* 75 (3), 199-211.
- Tillmann, K.-J. (1988). Comprehensive Schools and Traditional Education in the Federal Republic of Germany. *International Journal of Educational Research*, 12 (5), 471-479.
- Weinert, F.E. (1982). *Formen direkter Unterweisung*. Fernsehkolleg "Lehrerprobleme - Schülerprobleme", Heft 6A. Tübingen: Deutsches Institut für Fernstudien.
- Wright, B.D., Linacre, J.M. & Schulz M. (1990). *A User's Guide to BIGSCALE*. Chicago: MESA.