An Investigation on the Accuracy of Teachers' Judgments on Students' School Achievement

Constantinos Vouyoukas[1,*]

[1]School of Early Childhood Education, Faculty of Education Tower, Aristotle University of Thessaloniki, Thessaloniki 541 24, Greece

*Corresponding author: School of Early Childhood Education, Faculty of Education Tower, Aristotle University of Thessaloniki, Thessaloniki 541 24, Greece. Tel: 30-2310-991-299

Abstract

Teacher-based assessment of students' school achievement is a process in which collection of information on students' performance is needed to allow teachers to understand what and how students achieve. This process involves the development of teachers' judgments. This study investigated the predictive utility of an informal teacher-made achievement test on teachers' judgments to students' academic performance. The study involved 166 primary school students (mean age 11.7) and their respective 46 teachers. Main findings indicated that: 1. teachers' general judgments did not distinguish the difference between students with low achievement and underachievement, and 2. teachers' judgments with the use of an academic performance rating scale could differentiate low achieving from underachieving students. It is suggested that the accuracy of teachers' judgments on students' school achievement is increased when domain-specific and multi-point scales of students' academic performance and behavior are used.

Keywords: teacher-based assessment of school achievement, teacher-made school achievement test, academic performance rating scale, teachers' judgments, primary school students, Greek educational system

1. Introduction

1.1 Teacher-based Assessment of School Achievement

School achievement reflects students' level of learning at a given moment. It means that students can respond to school standards of their grade based on specific objectives and activities provided in the classroom environment. According to Glatthorn (1999), school achievement is a cycle, defined by the interactive relationship of standards-based curricula, performance assessment, assessment-driven instruction, and authentic learning. Therefore, assessment is an integral part of school achievement, namely the systematic collection of information. Teachers assess students' progress in relation to the curriculum objectives (e.g., by evaluating students' assignments and their response to instruction). Results of this assessment may be used for students' referrals for diagnosis. The assessment of students' school achievement may enable a more accurate evaluation on their eligibility for special education referral and the level of educational support they need to receive.

Teachers may not develop and construct tests that match the curriculum and state standards, so they are not very successful when comparing their students to grade norms (McLeskey, 1989). Especially in countries where there are no available standardized school achievement tests, it is impossible for teachers to know whether a student has made the number and type of errors that are normal and expected. Nevertheless, teachers are the only ones capable of matching tests to instruction and ranking students in their classrooms. To accomplish this, teachers use informal teacher-made tests and rating scales, which assess students' school achievement, academic performance and learning behavior. The use of these tests and scales involves the development of teachers' judgments, which are associated with students' placement at various achievement categories, such as low achievers, underachievers and high achievers (Ready & Wright, 2011). Even though defining and measuring 'average' is problematic (e.g., the use of cut off points in an achievement test -Fletcher & Miciak, 2019), still, in order to define underachievement there is a need to measure low achievement. Low achievement indicates that students are performing below average in terms of their grades and/or test scores (Shin et al., 1986). On the other hand, the difference that sets the underachievers apart from the low achievers is that underachievement is being documented through evidence of a significant discrepancy between intelligence test score and school achievement score, as opposed to a below average cognitive ability functioning and achievement (Flanagan et al., 2006). Because teachers cannot (and should not) apply intelligence tests to their students, they judge students' low achievement as unexpected when they believe that these students are 'smart', but due to cultural, environmental and motivational dysfunctional patters they perform poorly in the classroom. Also, because the basis of any conceptual model of learning disabilities is the issue of 'unexpected underachievement' -mainly because students with learning disabilities do not learn to read, write, and/or do arithmetic, despite the absence of conditions associated with low achievement (Fletcher, 2012)- it is reasonable to assume that if teachers believe that a low-performing student is learning disabled, then he/she may be considered as underachiever. An example of an underachiever may be a student with average and above average receptive and reasoning academic performance and below average expression and written language academic achievement.

Research findings indicate that, although teachers' judgments are positively associated with students' school achievement, they are accurate only for students with high school achievement, and not for low school achievers (Begeny et al., 2011). These research findings are associated with judgments' accuracy criteria. Literature review and meta-analysis research findings have shown that teachers' judgments, which are assessed by Likert-type scales focusing on different and several achievement domains, have more predictive value on students' actual school

achievement (Hoge & Coladarci, 1989; Sudkamp, Kaiser & Moller, 2012). Research has revealed that teachers' judgments accuracy on their students' academic performance is stronger when specific information and analytic questions are provided to the teachers before they develop their judgments to students' achievement in various school curriculum areas (Demaray & Elliott, 1998; Eckert et al., 2006; Hamilton & Shinn, 2003).

1.2 Teacher-based Assessment of Students' School Achievement in the Greek Educational Context

The current Greek statute on school education is the Law 1566/1985 and a series of decrees and government enactments based on it. Primary school teachers in Greece are required to follow the curriculum for all of their students and assess their school achievement in every teaching unit with the use of teacher-made school achievement tests. These tests evaluate the extent to which students have acquired the knowledge included in the relevant school textbooks, which use training exercises (Vlachou, 2018). School achievement is evaluated with marks. Since 1995 marks on the school certificates were only maintained in the last two school grades (Decree 121/1995) (5-10 and 'Scarcely good' for students receiving under 5) and those students who are systematically failing are being referred for evaluation in interdisciplinary centers in order to receive special education services. The referral is accompanied by a summary report of student's low achievement background history and teacher's assumptions on the causes of low achievement (Law 4547/2018). These causes are usually for most serious cases intellectual disabilities and, for milder cases, dyslexia or other learning difficulties, and environmental factors (e.g. low motivation, low social economical status -SES-, dysfunctional family backgrounds etc.). Recent research evidence indicates that teachers in Greece mainly attribute students' low school achievement in within-student problem attributions, that is a disability (e.g., dyslexia, hyperactivity or other specific developmental disorders and intellectual disabilities) (Michalopoulou & Vouyoukas, 2021). Teachers ' judgments on students' school achievement are usually created on the basis of a general student picture, rather on the actual evaluation results. Accordingly, it seems that teachers' prejudices are in place on students' intellectual abilities, which teachers cannot evaluate, but often express statements like "he/she has good receptive skills", "he/she has good emotional intelligence", "he/she is a clever kid but he does not learn" (Athanasiou, 2000). Research evidence by Mavrommatis (1996; 1997) have indicated that teachers' comments to students' training exercises were general and short, with little explanation of what students' strengths and mistakes were, and how improvements could be made or maintained; furthermore, teachers commented that often they had no clear idea of the criteria by which they assess students' school achievement and could not easily describe them.

Research indicates that a valid assessment of students' school achievement considers their pre-existing knowledge and the necessary skills to accomplish generalizations, including the cognitive abilities required by the students, which are important constituents of achievement as outcomes of learning (Peng & Kievit, 2020; Sternberg, 1998). In this way emphasis is put on assessment of competence (Gregorie, 1997), which includes assessing students' instructionally meaningful tasks of increasing levels of academic and cognitive complexity, the use of real-world applications, and significant commitments of students' time and effort (Palm, 2008). Therefore, a valid assessment of students' school achievement requires students to provide answers and produce results by activating knowledge and processes. The term 'knowledge' includes students' understanding and expression of information in various school subjects. The term 'processes' includes ways in which a student can express knowledge by using appropriate processing skills. It appears that the assessment of students' school achievement is fragmented if it does not include students' pre-existent knowledge and

processes. In assessing students' school achievement, it is not unusual that teachers in Greece put emphasis on the static knowledge and information provided in the school classroom, which results in neglecting students' thinking and reasoning (Kouloumparitsi & Kavouri, 1994; Vamvoukas & Kanakis, 1997).

2. Method

2.1 Study Rationale and Research Aim

The main assumption of this study is that primary school teachers in Greece do not assess in a valid way students' school achievement, but they rather focus on the static knowledge and neglect students' thinking processes. This results in the creation of general teachers' judgments on students' achievement, which may not differentiate low achieving from underachieving students and place these students under considerable risk, since their specific educational needs are not being met. Accordingly, if students' school achievement is not assessed in a valid way, it may result in unfair and stigmatized judgments and treatment of students whose low achievement may otherwise be improved with different classroom instruction by their teachers, and not necessarily by referral to special education services.

The aim of the research was to investigate the extent to which teacher-made school achievement test predicts teachers' judgments on students' academic performance. It was hypothesized that teachers' judgments with the use of an academic performance rating scale will be more accurate than teachers' general judgments in distinguishing low achieving from underachieving students.

2.2 Study Participants

Study participants were 166 students in the last elementary school grade from 34 primary schools in western Thessaloniki area, Northern Greece, and their respective schoolteachers (n=46). The average students' age was 11.7 years. 67.5% of the students were boys and 32.5% girls. 46.5% of the teachers were male and 53.5% female. The average teachers' age was 46.8 years.

2.3 Data Collection Tools

For the purpose of the research a teacher-made school achievement test was constructed. This test, in addition to the curriculum context, included assessment of skills associated with the knowledge that the student is expected to acquire at the school grade level he/she participates. Focus was placed on the assessment of achievement in written language because in Greek school education it is considered the main evaluation criterion for students' referral to special education services. For the construction of the test, the researcher collaborated with 15 elementary school teachers who taught in the last primary school grade and have acquired an MA degree in special and inclusive education. The curriculum context was considered, which in the last Greek elementary school grade focuses on text comprehension, writing, grammar and syntax. Six texts of progressive difficulty with narrative and explanatory formula, which included vocabulary based on the Greek school curriculum context, were selected. Written exercises using grammar and syntax rules, which had already been taught in school, were included in the grammar and syntax. Students were also asked to write an essay and argue for their views on the given essay (i.e, describe the characteristics of their best friend and write arguments supporting their choice). In this way students' understanding, expression, reasoning, and use of pre-existent knowledge of language rules and information in new situations and problems was assessed. In assessing written content generation, the quality, sequencing and

coherence of ideas, views and arguments was considered (Salvia et al., 2010). The teacher-made school achievement test was distributed to a pilot sample of 20 elementary school students attending the last grade. After the completion of the essays' marking by 2 of the teachers, 3 texts were rejected. Following a discussion between the teachers and the researcher, disputes were resolved over the grading of the rest of the texts with the percentage of agreement between the evaluators reaching 85%. Teacher-made school achievement scores were distributed in 4 quartiles with a maximum score of 100. The 1st quantile included scores from 0-24, the 2nd quantile included scores from 25-49, the 3rd quantile included scores from 50-74, and the 4th quantile included scores from 75-100. The first two quantiles considered low achieving students, the 3rd quantile included average performing students, and the 4th quantile considered high achieving students.

An academic performance rating scale was constructed by the researcher with the collaboration of the same 15 schoolteachers who contributed to the construction of the school achievement test. The rating scale was distributed to a pilot sample of 8 primary school teachers who answered the rating scale questions on the same pilot sample of 20 elementary school students who completed the teacher-made school achievement test. The academic performance rating scale included 4 domains (reception, expression, writing, reasoning) with an 'Often', 'Occasionally' and 'Seldom' Likert-type scale with scores 1, 2 and 3 respectively. The lowest score of the rating scale was 46 and the maximum was 138. The reception domain included 12 questions. The expression domain consisted of 12 questions. The writing domain included 10 questions. The reasoning domain consisted of 12 questions (see Appendix for full description). The application of the rating scale (α: 0.78), enabled the formation of three categories of students: low achievers (those who scored 46 and below), ambivalent students (those who scored 47-92) and underachievers (those who scored above 47 with high scores in reception and reasoning -above 24 respectively- and low scores in expression and in written language, below 24 respectively).

2.4 Procedure

Data collection was initiated in January 2020 and was completed in March 2020. Ethical approval was not required for conducting this study because of national laws. Written informed consent was obtained from all study participants, including teachers and students' parents. First, teachers were asked to indicate students with low school achievement on the basis of the general picture they had on their achievement in language subjects. Secondly, teachers were asked to attribute the causes of students' low achievement. Their answers provided their general judgments. If teachers indicated that the cause was mild intellectual disabilities, then these students were considered as low achievers; if they stated that the cause was dyslexia/learning disabilities and/or psychosocial difficulties, then these students were considered as underachievers; and if teachers indicated that the cause was unknown, then these students were considered as ambivalent (not being sure where the student is placed). The researcher administered the teacher-made school achievement test to the students on an individual basis. Finally, teachers completed the academic performance rating scale for every participant student. Their answers provided their rating scale judgments.

2.5 Analysis

Binary logistic regression equations were employed to address this study's research aims for the following reasons: 1. in this study, the outcomes were categorical variables with three (students with low school achievement, students with average school achievement and students with high school achievement) or four (school achievement quartiles) distinct values, 2. School achievement was selected as predictor based on the analysis of the Greek educational reality,

thus, it can be stated that the predictor included was evidence-based, 3. the categories in this analysis were mutually and collectively exhaustive, for example, the variable 'students with low school achievement' has two categories that are mutually and collectively, with or without low school achievement and 4. the number of participants in this study (n=166) was sufficient for logistic regression analysis (Mehta & Patel, 1996). The predictor variable was school achievement categorized in four quartiles as emerged by the teacher-made school achievement test assessment. The dependent variables were teachers' general judgments and rating scale judgments on students' school achievement. In this way, the binary regression analysis used in this research examined the extent to which teacher-made school achievement test predicts teachers' general and rating scale judgments.

3. Results

From table 1 we can see that, according to the school achievement test, the highest percentage of students (50.6%) was concentrated in the 3rd quantile (51-75), which consists of the average achievers. A considerable percentage of students (37.3%) was below the cut point of the class (50 and below), which consists of the low achievers. A small percentage (12%) of students was concentrated in the 4th quantile (76-100), which consists of the high achieving students. It is worth mentioning that all the research study participant students were considered as low achievers by their teachers' general judgments, an assumption which was not confirmed by the research findings.

Table 1. Frequencies and Percentages of Students' School Achievement Based on Teacher-Made Achievement Test (N=166)

1st quantile of achievement (0-24)	2nd quantile of achievement (25-49)	3rd quantile of achievement (50-74)	4th quantile of achievement (75-100)
4 (2.4%)	58 (34.9%)	84 (50.6%)	20 (12%)
	Low achievers	Average achievers	High achievers

Table 2. Frequencies and Percentages of Teachers' Judgments (*N*=166)

N=166	Teachers' general judgments	Teachers' judgments with rating scale
Low achievers	76 (45,8%)	68 (41%)
Underachievers	16 (9.6%)	38 (22.9%)
Ambivalent	74 (44,6%)	60 (36.1%)

Table 2 presents teachers' general and rating scale judgments for the same students before and after the distribution of the teacher-made school achievement test respectively. According to teachers' general judgments, 45.8% of students were considered low achievers, 9.6% were considered underachieves, while a high percentage (44.6%) of students was considered as ambivalent. According to teachers' judgments with the use of the rating scale, 41% of the students were considered to be low achievers, 22.9% were considered underachievers and 36.1% students were considered ambivalent. Noteworthy the increased percentage of the students considered underachievers based on teachers' judgments with the use of the rating scale as compared to the percentage drawn by teachers' general judgments (22.9% and 9.6% respectively).

Binary logistic regression analysis in Table 3 showed significant results in all students' groups with the use of the academic performance rating scale, which meant that: a. teachers' judgments on students' school achievement with the use of the rating scale could distinguish underachievers, low achievers and ambivalent students, and b. teacher-made school achievement test affected teachers' judgments when school achievement for all students' groups was evaluated with use of an analytical criterion, such as the academic performance rating scale. On the other hand, while significant results were shown when comparing low achievers and underachievers with ambivalent students with the use of general judgments, non-significant results emerged in comparing low achievers with underachievers. That is, teachers' general judgments were not affected by the teacher-made school achievement test on low achievers and underachievers, which meant that general teachers' judgments could not distinguish the low achieving from the underachieving students.

Table 3. Full Model Results per Outcome Group. Predictor: Teacher-made achievement test

Outcome group	B	SE	p	OR	Adjusted odds Ratios [95% CI]
Teachers' general judgments on low achievers and underachievers (*N*=92)	.139	.437	p=0.750 (NS)	1.15	[0.49, 2.70]
Teachers' general judgments on low achievers and ambivalent students (*N*=150)	1.108	.275	p=0.001	3.02	[1.76, 5,17]
Teachers' general judgments on underachievers and ambivalent students (*N*=90)	1.268	.451	p=0.005	3.55	[1.47, 8,59]
Teachers' judgments with rating scale on low achievers and underachievers (*N*=106)	1.133	.384	p=0.003	3.10	[1,46, 6,59
Teachers' judgments with rating scale on low achievers and ambivalent students (*N*=128)	1.687	.372	p=0.001	5.40	[2.61, 11,21]
Teachers' judgments with rating scale on underachievers and ambivalent students (*N*=98)	1.138	.308	p=0.001	3.12	[1.71, 5,70]

Note. B = coefficient. *SE* = standard error. *p*= significance level. *OR* = adjusted odds ratio for predictor variable.

4. Discussion

4.1 Points under Consideration

This study aimed to provide research evidence on the assessment of students' school achievement with the use of a teacher-made school achievement test and assumed that this test would differentially predict teachers' judgments on their students' academic performance. Results are discussed in the following three points.

First point: it should be noted that even though all the participant students in this research were considered as low achievers by their teachers, the administration of the teacher-made school achievement test differentiated them. Findings showed that approximately one third of the participant students could not meet the written language curriculum requirements and half of them experienced considerable difficulties in written language. These findings provide evidence suggesting that a school achievement test, which assesses students' knowledge and processes, allows for a sophisticated recording of students' school achievement. The study findings support the assumptions that Greek teachers' judgments on students' school achievement are usually created on the basis of a general student picture, rather on the

assessment of their actual achievement (Athanasiou, 2000; Mavrommatis, 1996; 1997).

Second point: looking at teachers' judgments on their students' school achievement, it is shown that almost half of the teachers' general and rating scale judgments attributed students' low achievement within-student problem factors (45.8% and 41% respectively). These findings are in accordance with research findings reported by Michalopoulou & Vouyoukas (2021) on teachers' attributions on students' low achievement. It appears that largely more underachieving students were identified with the use of the rating scale than when the general students' picture was considered by their teachers. This finding might suggest that the use of analytic, domain-specific and multi-point rating scales allows sophisticated judgments, which can differentiate the different needs of students with low achievement.

Third point: looking at the regression analysis results, it appears that general judgments, which are based on the general picture teachers have for their students, did not distinguish the difference between students with low achievement and underachievement. Because in Greece students' referrals are made on the basis of general teachers' judgments on students, a considerable number of referrals are noted. This results, especially in the language course, in students' stigmatization as underachievers and low achievers, mainly because achievement in written language is considered one of the main characteristics of the 'good' and 'bad' student. Furthermore, these students before they are referred for diagnosis, they should be evaluated by their teachers with the use of valid school achievement tests and analytic rating scales, which can evaluate, in addition to the static curriculum knowledge, students' preexisting knowledge and thinking processes. In this study it was found that teachers' judgments developed with the use of the rating scale, which analytically evaluates students' academic performance in different domains, could distinguish low achievers from underachievers more accurately. Current study findings are in accordance with research evidence on the significant predictive value accuracy of teachers' analytic judgments on student achievement (Demaray & Elliott, 1998; Eckert et al., 2006; Hamilton & Shinn, 2003; Hoge & Coladarci, 1989; Sudkamp et al., 2012). Findings from the second and the third point provide evidence which suggest that teachers' judgments on students' school achievement with the use of a rating scale are more accurate in distinguishing low achieving from underachieving students than general teachers' judgments are.

4.2 Conclusions and Educational Implications

It is worth mentioning that most of the participant students in this research, who study in mainstream classrooms, had low school achievement on the basis of the written language curriculum requirements. Moreover, approximately one third of them were under the classroom achievement cut point and should be referred for further evaluation because among them may be students at risk for developmental disabilities. Supporting teachers to develop analytic judgments on their students' school achievement accommodates their students' needs, especially those with low achievement. These students should receive support in order to improve their school achievement, which is crucial for their school and academic progress.

As Feinberg & Shapiro (2003) have argued, emphasis should be placed at the authentic assessment of students' school achievement and the development of accurate teachers' judgments to students' academic performance. In this way the accuracy of teachers' judgments on low achieving students and underachievers may be improved (Flynn & Rachbar, 1998). For teachers, the main criterion for evaluating students' academic difficulties is school achievement, which refers to the degree of response to the knowledge teachers provide to their students, rather than on the students' abilities. Consequently, for teachers two points may seem important when they evaluate students' low achievement and judge if it concerns underachievement or

not: first, to examine if they have authentically assessed their students' school achievement, and second, if they have associated their students' low school achievement with inadequate students' response to quality teaching instruction. In this way research on the assessment of students' school achievement may be placed in an educational-school context with measurable and changeable aspects of the classroom instructional environment.

This research has provided evidence, which demonstrated that teachers may not be able to provide an analytical student's school achievement profile with strengths and weaknesses in knowledge and processes. Accordingly, teachers' training on valid teacher-made school achievement assessment and analytic, domain-specific and multi-point scales of students' academic performance may be needed. This is particularly important in countries in which norm-referenced achievement tests are not in use. It is also recommended that in evaluating low achievers and ambivalent students' achievement, additional factors in the learning process should be considered, including motivation, psychosocial and sociocultural factors. In this way teachers may be more effective in the assessment of students' school achievement and in the identification of low achieving and underachieving students, thus contribute significantly to the referral procedure.

4.3 Future Research and Study Limitations

Further research may be required for evaluating students' instructional response with the use of curriculum-based assessment in main academic areas, as well as the use of Response to Instruction (RtI) approach. This research may contribute to the enrichment of children's psychoeducational evaluation, which could include a valid assessment of academic achievement as well as a reliable assessment of academic skills growth in relation to the quality of the instruction.

The following study limitations are noted: first, the study participants were not representative to enable generalizations of the findings; second, teacher-made school achievement test and academic performance rating scale are informal and not standardized tools, therefore their results must be treated with caution; third, school achievement test considered written language only and not reading comprehension; and fourth, the rating scale assessed teachers' judgments on students' academic performance at one point in time only, which might have resulted in some amount of recall bias.

References

Athanasiou, L. (2000). *The assessment of students' school achievement and school teaching*. Ioannina: University of Ioannina. (in Greek)

Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, *40*(1), 23-38. https://doi.org/10.1080/02796015.2011.12087726

Decree 121/1995. Complementary provisions on the assessment of primary school pupils. (in Greek)

Demaray, M.K & Elliott S.N (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8-24. https://doi.org/10.1037/h0088969

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the

correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, *43*(3), 247-265. https://doi.org/10.1002/pits.20147

Feinberg, A. B., & E. S. Shapiro. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research, 102*(6), 453-462. https://doi.org/10.3200/JOER.102.6.453-462

Flanagan, D. P., Ortiz, S. O., Alfonson, V. C., & Dynda, A. M. (2006). Integration of response to intervention and norm-referenced tests in learning disability identification: learning from the towel of Babel. *Psychology in the Schools*, *43*(7), 807-825. https://doi.org/10.1002/pits.20190

Fletcher, J. (2012). Classification and identification of learning disabilities. In B. Wong & D. Butler (Eds.), *Learning about learning disabilities* (pp. 1-25). MA: Elsevier. https://doi.org/10.1016/B978-0-12-388409-1.00001-1

Fletcher, J. M., & Miciak, J. (2019). *The identification of specific learning disabilities: a summary of research on best practices*. Texas Center for Learning Disabilities: University of Houston.

Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools*, *35*(2), 163-172. https://doi.org/10.1002/(SICI)1520-6807(199804)35:2<163::AID-PITS8>3.0.CO;2-Q

Glatthorn A. A. (1999) *Performance standards authentic learning*. NY: Eye of Education.

Gregorie, J. (1997). Diagnostic assessment of learning disabilities: From assessment of performance to assessment of competence. *European Journal of Psychological Assessment*, *13*(1), 10-20. https://doi.org/10.1027/1015-5759.13.1.10

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*(2), 228-240. https://doi.org/10.1080/02796015.2003.12086195

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, *59*(3), 297-313. https://doi.org/10.3102/00346543059003297

Kouloumparitsi, A., & Kavouri, P. (1994). The implementation of history curriculum in primary school: An empirical research. *Sigxroni Ekpaideysi*, *74*, 62-70. (in Greek)

Law 1566/1985. Structure and function of primary and secondary education and other provisions. (in Greek)

Law 4547/2018. Restructure of primary and secondary school support services. (in Greek)

Mavrommatis, Y. (1996) Classroom assessment in Greek primary schools. *The Curriculum Journal, 7*(2), 259-269. https://doi.org/10.1080/0958517960070209

Mavrommatis, Y. (1997) Understanding assessment in the classroom: Phases of the assessment process. The assessment episode. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 381-400. https://doi.org/10.1080/0969594970040305

McLeskey, J. (1989). The influence of level of discrepancy on the identification of students with learning disabilities. *Journal of Learning Disabilities*, *22*(7), 435-438. https://doi.org/10.1177/002221948902200708

Mehta, C., & Patel, R. (1996). *SPSS exact tests 7.0 for windows*. Chicago: Statistical Package

for the Social Sciences.

Michalopoulou, L. E., & Vouyoukas, C. (2021, August). *Teacher ratings on Roma student language and literacy performance*. Poster presentation at the Virtual American Psychological Association Annual Convention, Washington, DC.

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation*, *13*(4), 1-11. https://doi.org/10.7275/0qpc-ws45

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A biobirectional perspective. *Child Development Perspectives*, *14*(1), 15-20. https://doi.org/10.1111/cdep.12352

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335-360. https://doi.org/10.3102/0002831210374874

Salvia, J., Ysseldyke, J., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th ed.). Belmont: Wadsworth.

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labelled learning disabled and low achieving on measures of classroom performance. *Journal of Learning Disabilities*, *19*(9), 545-552. https://doi.org/10.1177/002221948601900906

Sternberg, B. J. (1998). Abilities and forms of developing expertise. *Educational Researcher*, *27*(3), 11-20. https://doi.org/10.3102/0013189X027003011

Sudkamp, A., Kaiser, J., & Moller, J. (2012). Teachers' judgments of students' academic achievement. *Journal of Educational Psychology*, *104*(3), 743-762. https://doi.org/10.1037/a0027627

Vamvoukas, M., & Kanakis, I. (1997). Teachers' attitudes towards multimedia teaching material. *Paidagogiki Aksiologisi. 25*, 61-86 (in Greek)

Vlachou, M. A. (2018). Classroom assessment practices in middle school science lessons: a study among Greek science teachers. *Cogent Education*, *5*(1). https://doi.org/10.1080/2331186X.2018.1455633

Appendix A

Student Academic Performance Rating Scale

<u>Reception</u>

1. It is difficult for him/her to recognize that sentences with different syntax have the same meaning (e.g. "The dog chased the cat" means the same as "The cat was chased by the dog").

2. He/she misunderstands simple written sentences and questions.

3. He/she misunderstands oral instructions.

4. He/she asks the teacher to repeat the instructions.

5. He/she does not fully understand what others are saying.

6. It is difficult for him/her to understand the meaning of polysyllabic words.

7. He/she does not understand the spoken word of others, but reads without difficulty.

8. He/she has difficulty in understanding sentences that are said at a fast pace, but he/she can understand them if they are repeated afterwards.

9. He/she finds it difficult to understand the content of a text while reading it.

10. It is difficult for him/her to distinguish two words that contain the same or different middle consonants.

11. He/she does not understand metaphors in speech.

12. It is difficult for him/her to remember multiple instructions.

Expression

1. He/she speaks slowly or simplistically.

2. He/she cannot repeat sentences that contain more than 5 words.

3. He/she uses immature syntax structure when speaking (e.g. does not use subordinate clauses).

4. It takes a while for him/her to recall words.

5. It is difficult for him/her to spontaneously express an opinion in the classroom.

6. He/she finds it difficult to express the meaning of a short text he/she has read.

7. He/she uses many abstract nouns (e.g. 'something', 'thing') and indefinite pronouns (e.g. 'this', 'that') in his/her speech.

8. He/she uses short or incomplete phrases.

9. He/she makes a lot of pronunciation mistakes.

10. He/she cannot pronounce common letter combinations (e.g. 'th', 'er', 'on', 'an').

11. He/she cannot repeat words and phrases.

12. He/she uses incomplete, fragmentary sentences.

Writing

1. He/she writes messily.

2. He/she writes slowly.

3. He/she makes many thematic and concluding spelling mistakes.

4. He/she has limited vocabulary.

5. His/her writing performance is insufficient (e.g. he/she writes short reports, few sentences and words).

6. He/she omits words in sentences.

7. He/she skips word endings.

8. He/she writes the words so misspelled, which makes it very difficult for someone to understand which word it is.

9. He/she writes fragmentary sentences.

10. He/she constantly writes small main sentences.

Reasoning

1. He/she does not address the main topic and focuses on minor details.

2. It is difficult for him/her to adapt to other children's different types of responses.

3. It takes a long time for him/her to answer simple questions.

4. He/she has inconsistency in reasoning and he/she develops irrational arguments.

5. It is difficult for him/her to learn abstract concepts.

6. It is difficult for him/her to organize, classify and form concepts.

7. It is difficult for him/her to evaluate thoughts.

8. It is difficult for him/her to make generalizations.

9. He/she cannot recognize cause-and-effect relations.

10. He/she cannot delve into generalizations

11. He/she cannot organize his/her thoughts into a coherent action plan.

12. He/she is rushing to draw conclusions.

Copyrights