Bolesław Niemierko
Akademia Marynarki Wojennej w Gdyni

EDUCATIONAL DIAGNOSTICS FOR CONTEMPORARY SCHOOL SYSTEMS. MEASURING AND ASSESSING GROWTH OF STUDENT HUMAN CAPITAL. PART III: SELECTED OUTCOMES

ABSTRACT

Educational practice abounds in *wishful thinking* which occurs when we believe that something is true because we want it were true. For example, in Poland we set very low cut-scores to external examinations (30%) and still believe that passing the examination means mastery of the subject.

Fortunately for educational diagnostics, assessment practice at schools involves emotional/motivational achievement what makes it a good predictor of future accomplishments at schools and on employment. Assessing growth of human capital do begin at general education schools in Poland and elsewhere (*US*), though in a latent, and partly illegal way, creating the second systems of grading. Frequently applied in lower tiers of education, *commented reports*, in which teachers describe student achievement in their own words, are more closely related to human capital assessment than the test-based grading used to be.

Both pedagogues and students are highly critical of standards of justice in the present achievement assessment practices what makes their consequential validity vulnerable to negative opinions. On the other hand, attempts to determine the worth of educational phenomena by large surveys in the shape of educational evaluation could be contaminated by various political factors. Widening the scope of educational diagnosis to cover the full range of human capital developmental aspects would encourage the better, more economically and ethically sound actions.

Keywords:
educational diagnostics, wishful thinking, achievement assessment, commented report, grading, consequential validity, educational evaluation.

WISHFUL THINKING IN EDUCATION

In the previous parts of the elaboration[1], assumptions, paradigms, methodology, and foci of educational diagnostics were considered with trust to assist education in preparing young generations to perform their school and vocational career tasks. Poor diagnostic procedures lead into uncertainty and, in consequence, to wishful thinking in education.

Wishful thinking occurs when we believe that something is true because we want it were true. Unfortunately, such thinking is commonplace in education where diagnostic procedures are vague, one-sided, and largely biased towards our hopes for student capacity and teaching effectiveness. Most applications of criterion-referenced measurement methods to the cognitive domain of student achievement are deeply disappointing. The numbers o students not reaching well justifiable and plausible performance criteria are greater than could be accepted.

Unforgettable experience of a massive flood of wishful thinking gained Polish educational diagnosticians in a representative survey of cognitive achievement carried out in year 1984[2]. Percentages of students who failed to match the basic standards at that time are presented in Table 1.

Table 1. The outcomes of criterion-referenced tests in Mother Tongue and Mathematics in a national achievement survey in primary and secondary schools in Poland (1984)

Subject	School and grade	Item form*)	Curricular area	Passing score	Failure rate
Mother Tongue	Elementary 4	MC + O	*Minimum competence*	60%	25%
	Elementary 4	MC	*Reading comprehension*	65%	30%
	Elementary 4	MC	*Grammar and spelling*	60%	48%
	Elementary 8	MC + O	*Language knowledge*	45%	49%
	Elementary 8	MC + O	*Reading epic stories*	35%	51%
	Elementary 8	O	*Writing reports*	70%	53%
	Secondary 4	MC + O	*Language knowledge*	35%	37%

[1] B. Niemierko, *Educational diagnostics for contemporary school systems. Measuring and assessing growth of student human capital. Part I: Main concepts and the scope*, *Colloquium*, 2012, 1; B. Niemierko, *Educational diagnostics for contemporary school systems. Measuring and assessing growth of student human capital. Part II: Methodology and rules*, *Colloquium*, 2012, 2.

[2] B. Niemierko, *Ogólnopolskie badania osiągnięć uczniów, nauczycieli i szkół. Synteza badań* (National study of student, teacher, and school achievement. Research synthesis], Centrum Doskonalenia Nauczycieli, Warszawa 1990.

	Secondary 4	MC	*Poetry reception*	(lack)	59%
	Secondary 4	MC	*Reading contemporary poetry*	35%	31%
	Secondary 4	MC + O	*Self-education skills*	60%	62%
Ma-	Elementary 4	O	*Minimum competence*	75%	52%
thema-	Elementary 4	O	*Multilevel test*	75%	69%
tics	Elementary 8	O	*Minimum competence*	75%	73%
	Elementary 8	MC	*Multilevel test*	75%	74%
	Secondary 4	MC	*Multilevel test*	75%	62%

*) MC – multiple choice, O – open question

Passing scores for the tests gathered in Table 1 were much debated and carefully tried out by the eminent subject-matter specialists in our country, so the scores should be assumed to be adequate to the curricular standards and test difficulties. However, almost a half of the students did not reach the standards in Mother Tongue and about two third of students failed in Mathematics. Clearly, curricula and passing scores as well as test items represented wishful thinking of pedagogues rather than educational reality.

Fifteen years later, when external examination system was introduced into primary and secondary schools in Poland, the lesson learned in previous achievement surveys prevented educational decision makers from determining any cut scores for elementary and junior-high schools. All students are graduated from elementary school and from gymnasium, regardless of their examination score. Solely the senior-high school (lyceum) graduates with selective examination („matura"), nevertheless its cutting score is extremely low (30%).

Is it justifiable? Table 2 reveals the possible dramatic consequences of enforcing some passing scores upon the elementary school finals[3].

[3] A. Brożek, D. Grabowska, H. Jędrasik, J. Walczak, *Osiągnięcia uczniów kończących szkołę podstawową w roku 2007. Sprawozdanie ze sprawdzianu 2007* [Achievement of students graduated from elementary school in 2007. A report on external examination 2007], Centralna Komisja Egzaminacyjna, Warszawa 2007.

Table 2. The anticipated consequences of setting performance standards on the obligatory external examination in elementary schools in Poland (2007)

Passing score	Raw score required (out of 40 points)	Percentage of failure	Denied diplomas in thousands
70%	28	49,6	220
60%	24	33,8	150
50%	20	20,4	91
40%	16	10,2	45
30%	12	3,8	17
20%	8	0,9	4

Only 30% passing score could be reasonably applied to high-stakes examination in Poland and this is exactly the „rule of thumb" accepted for high-school examination („matura"), where curricula are wider, tests are more difficult, and wishful thinking is even more powerful than at the elementary level.

Nobody is satisfied with such a low performance level. Test scores in lower sections of achievement measurement scales have limited *content meaning*, they do not tell us what a student can do and what he/she cannot do in the selected curricular area when merely one third of test items are properly solved. However, the rigid academic tradition of European education prevents educational authorities from more realistic approach to designing and assessing the cognitive domain of human capital they foster.

ASSESSMENT PRACTICE AT SCHOOLS

Although in most countries subject-matter grading is formally confined to the cognitive domain, actually it is rather „a hodge-podge of student attitude, effort, and achievement"[4]. Motivational aspect plays an important and in many cases a leading role in teacher assessment of student learning. It may be seen in *commented educational reports* in which teachers describe student achievement in their own words rather than by means of a letter-grade or another formalized grading system. This kind of reports has been

[4] S. M. Brookhart, *Grading practices and validity*. *Educational Measurement: Issues and Practice*, 1991, 1.

strongly criticized both in Europe[5] and in the United States[6] but commonly used as essential in lower grades of many elementary schools and as an auxiliary assessment method in higher grades.

Here are some excerpts from a commented educational report in the form of a letter addressed to Margaret, a student of Grade 1 in an elementary school in Gdańsk[7]:

Maggie, you all-rounder,

You are a wise and smart girl.

Your reading is excellent, fluent, and expressive. Your writing from memory and from hearing is faultless. Though you write with shapeless letters and you often carelessly go out of the line. You tell stories willingly and well. Your conclusions are correct (…).

As you see, you are a dream schoolchild, and I guess that a dream daughter as well. You may be characterized in superlatives. However, Maggie, your work, your effort put into the final effect must also be assessed.

I think that your work at writing accuracy was too little. In your case, reading did not require any practice because you could already read well when you entered the school (…).

I was sure that you first of all children would finish the exercise <I count and paint>. However when other children started the task and evened you out, you stopped your work. Would only the desire to show off and not your wish for improvement and better knowledge guide your learning behavior? I do not suspect you of being driven by such vanity but it looks like that.

You enjoy attracting other children's attention. It is good that with your knowledge and eloquence and not with tantrum or whimsicality. Anyway, as you can see, your classmates stay a bit away from you. (…) Learn also how to listen to the others, accept the views of your colleagues, praise them for something and do not emphasize that you do it better (…).

I wish you to stand firm on the position of a model schoolchild and daughter. Your parents are certainly proud of such great first-grader. I wish them to be always happy about you. And to you – a successful rest in the mountains and getting great strength to work in the second semester.

---

[5] P. Krope, *Ocena opisowa w pedagogice jako źródło nieporozumień* [Educational commented report as a source of misunderstanding], [in:] B. Niemierko (ed.), *Diagnostyka edukacyjna*, Wyd. UG, Gdańsk 1994.

[6] S. M. Brookhart, *Grading*, Pearson, Upper Saddle River 2004.

[7] B. Niemierko, *Ocenianie szkolne bez tajemnic* [Grading without mystery], WSiP, Warszawa 2002, p. 194-197.

Main comments to the letter are following:

1. The letter goes far beyond the cognitive domain. It concerns also motivational domain (effort), moral domain (cooperation with classmates), and physical domain (handwriting). All the areas of Maggie's human capital development are subjects of teacher observation and evaluation.

2. The letter is personal, private, non-schematic. However, it seems to be addressed to Maggie's parents rather than to her since its vocabulary („all-rounder", „superlatives", „vanity", eloquence", „tantrum", „whimsicality") is certainly too difficult to elementary school first-grader.

3. Proportion of the number of sentences containing positive and negative information about Maggie's behavior is in the excerpt 10 : 8, and in the whole document 21 : 18. The two kinds of characteristics create a mixture of different components which may be difficult to synthesize for parents.

4. Superlatives like „excellent reading", „faultless writing", „dream schoolchild", „great first-grader" would not be probably proved by measurement procedures. Informal educational diagnostics often takes extreme views on a student's behavior (*halo effect*) while standardized diagnostics keeps to the mean of the feature.

5. Small incidents like breaking work on „I count and paint" exercise, quoted as the facts supporting general observations, should be tactfully concluded on the spot in the classroom or immediately after the lesson and not be included into a semester certificate. They do not deserve fixing on the paper and being provided to the child's parents.

Because of significant differences in focus and in procedure, school grades and standardized tests bring results widely divergent from each other. It can be seen in the outcomes of Educational Testing Service study of 8.5 thousand high-school graduates in the United States[8], presented in Table 3.

Table 3. Correlation coefficients of selected variables with school grades and standardized test scores

Variable	Correlation with school grades	Correlation with test scores
1. Educational motivation in teacher opinion	0,63	0,45
2. Work completed in teacher opinion	0,61	0,33
4. Class behavior in a teacher opinion	0,51	0,35
5. Parental socio-economic status (*SES*)	0,35	0,48

[8] W. W. Willingham, J. M. Pollack, C. Lewis, *Grades and test scores: Accounting for observed differences*, Journal of Educational Measurement, 2002, 1.

6. Student educational plans	0,35	0,33
7. School attendance and lack of delays	0,33	0,19
9. Student self-esteem	0,29	0,28
10. Peer studiousness	0,29	0,25

Two conclusions may be drawn from the data contrasted in Table 3:

1. Teacher estimation of a student's motivational achievement makes a fairly good predictor of her/his school grades but much weaker predictor of test scores.

2. While the teacher opinion on student attitude and effort strongly influences school grades it is less susceptible to the family *SES* than external examination scores. No doubt that teachers compensate for social inequalities while test scores are rigorous in keeping to the cognitive dimension of evaluation.

The fairly naïve question would be why teachers contaminate their subject-matter assessment with such circumstances like student effort, homework, good behavior, school attendance? Is it simply a class management strategy or an inherent quality of their educational work? Willingham and his coworkers content themselves with the following statement:

Both grades and test scores play an important role in high-stakes educational decisions. Tests are often used because of uncertainty about the meaning of grades, yet grades are used to evaluate the validity and fairness of tests. Grades and tests provide this mutual support because it is commonly assumed that they do or should measure much the same thing. Yet the two measures often yield somewhat different results. (…) Due to their distinguishing characteristics, grades and tests have different strengths that tend to be complementary. Common advice that the two measures should be used together where possible is well founded[9]:

The differences between examination scores and school grades consistently move along the levels of education and as a result of that prediction validity of high-school grades for college grades may be higher than prediction validity of the most sophisticated batteries of achievement tests (*SAT*, *ACT*).

Measurement theorists complain:

And yet, study after study shows that the predictive ability of college admissions tests and high school grade point average (or class rank) is roughly the same, and somewhat redundant. One wonders how this could be. How

[9] Herein, p. 31-32.

can crack teams of the world's best psychometricians be consistently battled to a draw in predicting college success by idiosyncratic collection of high school teachers? The answer to the question is embedded somewhere in the similarity in the processes that go into making a grade point average at the high school level and making one at the college level. As first described by Brookhart[10], and then documented by Cross and Frary[11], grading is a hodge-podge of attitude, effort, and achievement at the middle and high school levels. Some recent work (…) indicates that the situation does not change much at the college level[12].

We do not have enough reliable research to ascertain whether the same „processes similarity" extends as far as job market and job success. However, we believe that converting „the hodge-podge of student attitude, effort, and achievement" both into a set of standardized measurement scales and into a legal directives on motivational, moral, experiential, and physical assessment in classroom education is favorable to manpower market and, in the further perspective, to the country economy.

TWO STANDARDS OF JUSTICE IN GRADING

The concept of two standards of justice in grading was originated in Susan Brookhart's paper *Teachers' grading practices. Meaning and Values* (1993). She wrote:

There is a double standard of just deserts: An average or about average student gets „what (s)he earns", while a below-average student gets „a break" if there is any way to justify it. The difference is how the teacher perceives the student and reflects the teacher's advocacy function. (…) „I could not fail a student who was trying" because a student who „works hard" does not „deserve" to fail. (…) Recommended grading practices, suggesting no compromises, are of limited help to teachers on this issue. The study's results suggest that teachers mix the roles of judge and advocate differently for students of different ability, and this in itself

[10] S. M. Brookhart, *Grading practices and validity*, Educational Measurement: Issues and Practice, 1991, 1.

[11] L. H. Cross, R. B. Frary, *Hodge-podge grading: Endorsed by students and teachers alike*, Applied Measurement in Education, 1999, 1.

[12] J. K. Smith, *Reconsidering reliability in classroom assessment and grading*, Educational Measurement: Issues and Practice, 2003, 4.

is a value-laden act[13].

These general statements were supported by the views of 84 teachers who analyzed a number of anecdotal stories on assessment problems. The research outcomes are summarized in Table 4.

Table 4. Opinions of American teachers on achievement assessment at schools

Situation	Decision about the final grade (in percentages)		
	No change	Raise	Lower
1. Good achievement but below the student's ability	80	4	16
2. Despite hard work a student does not reach standard	6	94	–
3. Satisfactory achievement without making effort	95	–	5
4. Very good on the test but lack of homework	51	1	48
5. Satisfactory on the test but lack of homework	86	1	13
6. A weak student begins to get better	29	71	–
7. A good student approaches the grade „very good"	73	27	–
Average	60	28	12

Table 4 shows that American teachers are inclined to raise the grade to weak students who work hard (situations 2 and 6) and even to forgive them lacking homework (situations 5) but at the same time they present unyielding attitude (situations 4 and 7), and even rigidity (situations 1 and 3) toward those who they consider to be good students. Teachers who have passed educational measurement courses (n = 40) do not differ in opinions from those who have not got such special qualifications (n = 44). Both groups pointed out to the student's effort seen as the greatest value in education and to expected motivational effects of permissive policy on grading weak students' achievement.

Brookhart's concept of two standards of justice in grading was further elaborated in Poland[14]. *The second system of grading*, creating „the grey ar-

[13] S. M. Brookhart, *Teachers' grading practices: Meaning and values,* Journal of Educational Measurement, 1993, 2, p. 140-141.

[14] B. Niemierko, *Ocenianie szkolne bez tajemnic* [Grading without mystery], WSiP, Warszawa 2002. Chapter XIV.

ea" of assessment policy, is assumed to have the following characteristics:

1. It is a cross-national and inter-systemic phenomenon. It probably concerns every non-selective primary and secondary school as well as some higher education institutions.

2. It has its roots in the context intervention into educational processes. Not all the students have favorable external and internal conditions to intensive academic development. In too numerous cases the conditions are far from the assumptions of curriculum makers.

3. When it concerns students without formal psychological statement of the need of special treatment it is not entirely legal. On account of that, schools do not officially declare that they reduce educational standards and employ more lenient grading system.

4. It operates mainly in the lower part of grading scale. Teachers defend the upper part of the scale from *grade inflation* but their efforts to keep up high instructional standards are not always successful.

5. Both individual and social *feedback* of the second system is generally positive. Failures are inefficient incentives to learning and society does not profit from educational drop-outs. Successes are more productive.

6. Unfortunately, the second system makes actual grading procedures complex and confusing. Because of double structure, rules of the game are changeable and unclear. A student who is promoted by favor may lose a part of his/her learning motivation and self-esteem.

7. Four strategies may be applied by teachers to deal with the problems of grading weak students' achievement:

a. Open friendly strategy: *allowances for slower development*. Nobody should be ashamed to benefit from the reduced tariff and nobody is allowed to stop the effort to achieve standards. The teacher devotes much time to support slower learners.

b. Open unfriendly strategy: *fighting outsiders*. Every student realizes his/her weaknesses and delays but some students are evidently at risk of being a failure. The teacher repeatedly declares that her mission in education is to force the latter group to greater effort.

c. Hidden friendly strategy: *discreet empathy with slow learners.* The second system of grading does not officially exist but the teacher secretly helps weaker students on tests and examinations. Every student can expect understanding his/her problems if only demonstrates positive attitude toward learning.

d. Hidden unfriendly strategy: *covert aggression*. At the beginning

of the course everything seems to be in order but afterwards whenever appears class-management problem, the teacher attacks students with hard questioning on subject-matter details. As a result, weaker students learn how to avoid punishment rather than how to accelerate their progress.

Recently, Grażyna Szyling carried out a representative study of teachers (n = 120) and students (n = 480) of the Polish junior-high schools on the second system of grading[15]. The main outcomes of her research were following:

1. About a half of grading problem situations teachers resolve by *occasional lowering achievement standards.*
2. On *low-stake examinations* teachers are more demanding than on *high-stake examinations*. Their permissiveness grows alongside with the importance of examinations.
3. The second system of grading is applied to the whole range of achievement levels. There is no significant correlation between the levels and decisions about lowering standards.
4. Nor occasional lowering achievement standards correlates with school achievement levels measured by mean scores on country-wide external examinations.
5. An overall hierarchy of factors that make teachers decide to lower the standards cannot be identified. In everyday situation, a student's *effort* and *diligence* are the most important but at the end of school year the *personal and social consequences* of the decision count more.
6. Students are „generally convinced that everybody should be assessed exactly the same" but they approve lowering demands on those colleagues who heavily struggle for better learning outcomes.
7. Students gain knowledge about occasional lowering standards from their colleagues (66%) and by free observation (44%) rather than from subject-matter teachers (26%) or the homeroom teacher (7%).

As it was mentioned in the previous section of the paper, the high-school grade-point average in the United States is usually found to be a stronger predictor of college achievement than are the Scholastic Achievement Test scores. However, both the grade average and the test scores overestimate first-grade college grade-point average of low *SES* and minority (African-American and Latino) students, *i.e.* these students are predicted to

[15] G. Szyling, *Nauczycielskie praktyki oceniania poza standardami* [Teachers' practices of assessment out of standards], Impuls, Kraków 2011.

earn higher college grades than they actually do[16]. It is hypothesized that living troubles, financial problems as well as anxieties, low aspirations, or negative attitudes may interfere with their academic success. College professors are less able than high-school teachers to compensate for the scarcity of prerequisites for success.

The existence of second system of grading was also tracked down by the Polish sociologists of education[17]. They found that youth from low-*SES* families received on gymnasium examinations lower score than could be expected on the basis of their school grades. They explain the phenomenon by lesser *cultural capital* of the families but it could also be ascribed to the support given by the teachers who know the students' home learning conditions. There is no such special treatment of the low-*SES* students in external examinations. The researchers conclude their findings with such a bitter statement: „It is a paradox that the youth who was not treated as potential beneficiary from the reform, because they were already privileged, gained the most of introducing external examinations."

THE ETHICS OF EDUCATIONAL DIAGNOSIS

Samuel Messick's concept of *consequential validity*[18] has drawn the interest of educational measurers to the problem of their responsibility for far-reaching effects of their work. He distinguished four facets of measurement validity as presented in Table 5.

Table 5. Facets of Validity[19]

	TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity	Construct validity + Relevance/utility
CONSEQUENTIAL BASIS	Value implications	Social consequences

[16] R. Zwick, I. Himelfarb, *The effect of high school socioeconomic status on the predictive validity of SAT scores and high – school grade point average*, „Journal of Educational Measurement", 2011, 2, 101-121.

[17] J. Domalewski, P. Mikiewicz, *Młodzież w zreformowanym systemie szkolnym* [Youth in the reformed school system], IRWiR PAN, Toruń 2004.

[18] S. Messick, *Test validity and the ethics of assessment*, „American Psychologist", 1980, 35, 1012-1027.

[19] S. Messick, *Validity*, [in:] R. L. Linn (ed.), *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989.

Values, such as self-expression (assertiveness) of students and, on the other hand, their self-control, may both strengthen examination feedback on their performance and contaminate assessment with prejudice or perverse interpretations. Messick admits that self-expression and self control as well as many other personal and social traits are „open to conflicting value interpretations"[20] but warns us against ignoring the prevailing scientific theories, social ideologies, cultural idiosyncrasies, and personal inclinations. These circumstances may alter the meaning of measurement scores to particular groups of educational stakeholders.

There are intended and unintended *social consequences* of test interpretation and use. At the end of the previous part of the paper some negative consequences of external examination for low-*SES* family students were considered. *Side-effects* like backwash-effect, the test influence upon the content and method of learning long before examination („learning to the test"), and *by-products* of testing like test-wiseness instead of subject-matter knowledge are well known defects of contemporary cognitive assessment methods[21].

The most far-reaching social consequences of high-stake achievement assessment reduced to the cognitive domain are pertinent to emotional and moral domains of human capital development. According to Maria Groenwald, when examination success is the only true value in education – „(…) all ways leading to it become acceptable. This way an unofficial permission is given to cheating and lying throughout examination; unofficial because inconsistent with examination procedure, nevertheless allowed while not detected. Symptomatic of this attitude is ignoring the examination dishonesty by parents, teachers, society, mass media; though they not participate in this deception, they do not fight it and only sporadically pretend they do"[22].

The author of this harsh criticism comes to the conclusion that examination evil prevails over examination good even if cheating and lying are excluded. *Examination evil* involves „hidden curriculum, depersonalization of students and teachers, their humiliation, wish to passively conform to examinations and to follow them in teaching, *i.e.* teaching by the tests and to the tests, fierce competition, the pain of being stigmatized and excluded"[23]. *Examination good* involves only the process of intensive learning on account of expected examinations and the confidence in teachers' ability to prepare

[20] Herein, p. 60.

[21] B. Hoffman, *The tyranny of testing*, Crowell Collier, New York 1962.

[22] M. Groenwald, *Etyczne aspekty egzaminów szkolnych* [Ethical aspects of school examinations], Wydawnictwo UG, Gdańsk 2011, p. 30.

[23] Herein, p. 172.

students for passing them smoothly[24].

Is classroom, presumably *formative*[25] assessment more ethically valuable? A lot of research indicate negative answer to the question, especially when student opinion is concerned. Table 6 presents the outcomes of one of the studies. One hundred of seven – and eight-graders were confronted with several fictitious stories and asked: 1. „Was it fair?", and 2. „Could it happen in your classroom?"[26].

Table 6. Student opinions on classroom achievement assessment

The case	Justice [(-1) – (+1)]	Occurrence [percentage]
1. *Martin always gives his opinion about the book he read and is able to justify his view. The teacher gives him an A for independent thinking.*	+0.82	69
2. *Mary receives As because she is the fastest in mathematical problem solving.*	+0.15	55
3. *Julie receives As because she always adds some interesting details that are missing in a textbook.*	+0.63	44
4. *Helen on her own free will solves mathematical problems which were not assigned, so the teacher gives her As.*	+0.20	49
5. *Kathy gets a C because she offers an answer to the teacher's question although she knows very little about the subject.*	-0.16	59
6. *Matt solved the problem by himself but did not get a grade because he had not follow the method shown on the lesson.*	-0.79	54
7. *Patrick is considered a good pupil by the teacher and she always tries to give him an A though he seldom deserves it.*	-0.80	76
8. *The teacher does not like Alice and always tries to prove that she does not deserve any more than a D.*	-0.88	66
9. *Tom gets good grades for homework cribbed from his colleagues.*	-0.72	83
10. *The teacher gave some students very low grades because she was in a bad mood one of these days.*	-0.88	82

[24] Herein, p. 170.

[25] S. M. Brookhart, *Editorial. Special issue: The validity of formative and interim assessment*, Educational Measurement: Issues and Practice, 2009, 1.

[26] J. Denc, *Nauczycielskie modele oceniania osiągnięć w szkole podstawowej* [Teacher models of achievement assessment in primary schools], Unpublished master's dissertation, University of Gdansk, 1994; B. Niemierko, *Ocenianie szkolne bez tajemnic* [Grading without mystery], WSiP, Warszawa 2002, p. 235.

We can see in Table 6 that all the stories match the students' experiences (occurrence possibility of 57% on average) but the cases of evident injustice (from 6 to 10), strongly condemned by the students (-0,81 on average), are even more frequent (72% on average). While „most teachers dislike evaluating their students and giving grades"[27], most students certainly dislike assessment procedures the teachers apply in classrooms.

DIAGNOSIS AS THE KEY TO EVALUATION

Educational evaluation means gathering information about context, progress, and outcomes of learning in order to estimate its merits and faults and to recommend appropriate institutional decisions.

Educational evaluators attempt to determine the worth of educational phenomena: student achievement, teaching methods, school management, system efficiency[28]. They make use of diverse evidence, ranging from *anecdotal* stories and *folklore* beliefs to *descriptive* records and *research* (experimental) studies[29]. The most important evidence comes from educational diagnoses, both informal and standardized. Diagnostics and evaluation (the term „evaluatics" could be coined by analogy) are closely related as it is shown in Figure 1[30].

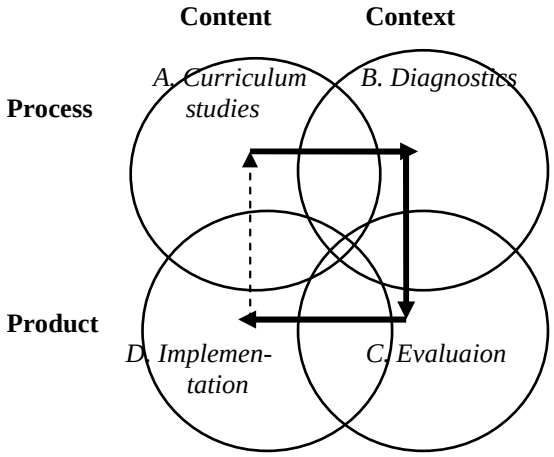
[27] L. A. Shepard, *Classroom assessment*, [in:] Brennan (ed.) *Educational measurement. Fourth edition*, American Council on Education – Praeger, Westport 2006, p. 637.

[28] W. J. Popham, *Educational evaluation*, Prentice-Hall, Englewood Cliffs, 1975, p. 175.

[29] A. W. Astin, R. J. Panos, *The evaluation of educational programs*, [in:] R. L. Thorndike (ed.), *Educational measurement. Second edition*, American Council on Education, Washington 1971.

[30] B. Niemierko, *Pomiar wyników kształcenia* [The measurement of teaching outcomes], WSiP, Warszawa 1999, p. 290.

Figure 1. Four types of applied research in education



Curriculum studies are content – and process-oriented. They answer two questions: „what to teach?" and „how to teach that effectively?". *Diagnostics* will make sure that so directed learning runs adequately to individual needs and circumstances, *i.e.* context aspects are considered crucial. *Evaluation* is product – and context-oriented. It focuses on the systemic value of student, teacher, and school achievement. When conclusions are positive, the now evaluated processes may be further implemented and they may lay the foundations for successive curriculum studies.

Evaluation theorists warn of reducing diagnostic basis for evaluative reports to student achievement testing[31]. „Clinicians and counselors generally use measuring devices to help make decisions about individuals" – remarked James Popham[32]. Occasionally, educational evaluators may also use them to support administrative decisions but they are

[31] L. Korporowicz, *Refleksja jako działanie. O stałej potrzebie przełamywania redukcji w pojmowaniu badań ewaluacyjnych w edukacji* [On the permanent need of breaking reductionism in understanding evaluation studies in education], [in:] B. Niemierko, M. K. Szmigel (eds.), *Regionalne i lokalne diagnozy edukacyjne*, PTDE, Kraków 2012.

[32] W. J. Popham, *Educational evaluation*, Prentice-Hall, Englewood Cliffs, 1975, p. 172.

not obliged to.

An American nationwide survey on public educational needs[33] showed that, at least at the elementary school level, affective goals (self esteem, socialization, need achievement, school orientation) are far more important (4.29 in 1 – 5 scale) than cognitive goals (reasoning, creativity, memory: 3.38 points), and physical education-health safety scored higher (3.59) than intellectual achievement, though not as high as reading (comprehension and interpretation: 3.91).

Political context plays a substantive role in educational evaluation. „Worth determinations alone will not be the only factors involved in educational decisions" – emphasize Popham[34]. And he illustrates the statement as follows:

Suppose that a harshly negative evaluation of a innovative program in university-level instruction for minority students would results in the abolition of the program and, thereby, at least for the time being, the elimination of the university's most visible effort to assist minority students. In addition, the abolition of the program would result in the dismissal of 15–20 staff members. Most of whom are representatives of minorities themselves. Now any evaluator who expects the involved decision-makers to instantaneously adopt the „shut it down" recommendation inherent in the adverse evaluation report is an evaluator in need of some seasoning[35].

Henry Brickell, then the head of an educational evaluation centers, summarized his experience in dealing with political decision-makers with a statement that sampling, methods, research designs, interpretations, conclusions, reporting styles, and, above all, consequences of evaluation are dependent on educational authorities[36]. „It is almost inevitable that an evaluation has a political dimension to it" added Colin Robson[37].

Teachers may not approve measurement-based supervision of their

[33] R. Hoepfner, P. A. Bradley, W. J. Doherty, *National priorities for elementary education,* University of California, Center for the Study of Evaluation, Los Angeles 1973.

[34] W. J. Popham, *Educational...*, dz. cyt., p. 302.

[35] Herein, p. 300.

[36] H. Brickell, *The influence of external political factors on the role and methodology of evaluation,* Evaluation Comment, 1976, 2, p. 5.

[37] C. Robson, *Real world research. A resource for social scientists and practitioner-researchers,* Blackwell, Oxford 1993, p. 183.

work[38] and teachers' associations used to question „negative evaluation of any of their members, even the manifestly incompetent, feeling that the dismissal of one teacher might trigger the future dismissal of many"[39].

Entangled in complicated political dependencies, educational evaluation suffers from many deformations which may be termed *pseudo-evaluations* (Suchman, 1967; Patton, 1981; Korporowicz, 2012)[40]. In *liberal* educational systems appear „posture", „postponement". „eyewash", „mock", „quick-and-dirty", and „amusing" evaluations. In *authoritarian* systems we may expect „weighty", „compliant", „fragmentary", and „submarine" evaluations. Even in *democratic* educational systems „good-wish", „whitewash", „personality-focused", and „guesstimate" evaluations are frequent.

In general, educational evaluation can make a good use and a deliberately deceiving use of diagnostic data.

BIBLIOGRAFIA

[1] Astin A. W., Panos R. J., *The evaluation of educational programs*, [in:] R. L. Thorndike (ed.), *Educational measurement. Second edition*, American Council on Education, Washington 1971.

[2] Brickell H., *The influence of external political factors on the role and methodology of evaluation*, „Evaluation Comment", 1976, 2.

[3] Brookhart S. M., *Grading practices and validity.* „Educational Measurement: Issues and Practice", 1991, 1.

[4] Brookhart S. M., *Teachers' grading practices: Meaning and values*, „Journal of Educational Measurement", 1993, 2.

[5] Brookhart S. M., *Grading*, Pearson, Upper Saddle River 2004.

[6] Brookhart S. M., *Editorial. Special Issue: The Validity of Formative and Interim Assessment,* „Educational Measurement: Issues and Practice", 2009, 1.

[38] D. A. Goslin, *Teachers and testing*, Russell Sage Foundation, New York 1967.

[39] W. J. Popham, *Educational…*, dz. cyt., p. 301.

[40] E. A. Suchman, *Evaluative research: principles in public service and action programs*, Sage, New York 1967; M. Q. Patton, *Creative evaluation*, Sage, London 1981; L. Korporowicz, *Refleksja…*, dz. cyt.

[7] Brożek A., Grabowska D., Jędrasik H., Walczak J., *Osiągnięcia uczniów kończących szkołę podstawową w roku 2007. Sprawozdanie ze sprawdzianu 2007* [Achievement of students graduated from elementary school in 2007. A report on external examination 2007], Centralna Komisja Egzaminacyjna, Warszawa 2007.

[8] Cross L. H., Frary R. B., *Hodge – podge grading: Endorsed by students and teachers alike*, „Applied Measurement in Education", 1999, 1.

[9] Denc J., *Nauczycielskie modele oceniania osiągnięć w szkole podstawowej* [Teacher models of achievement assessment in primary schools], Unpublished master's dissertation, Uniwersity of Gdansk 1994.

[10] Domalewski J., Mikiewicz P., *Młodzież w zreformowanym systemie szkolnym* [Youth in the Reformed School System], IRWiR PAN, Toruń 2004.

[11] Goslin D. A., *Teachers and testing*, Russell Sage Foundation, New York 1967.

[12] Groenwald M., *Etyczne aspekty egzaminów szkolnych* [Ethical aspects of school examinations], Wydawnictwo UG, Gdańsk 2011.

[13] Hoepfner R., Bradley P. A., Doherty W. J., *National priorities for elementary education*, University of California, Center for the Study of Evaluation. Los Angeles 1973.

[14] Hoffman B., *The tyranny of testing*, Crowell Collier, New York 1962.

[15] Korporowicz L., *Refleksja jako działanie. O stałej potrzebie przełamywania redukcji w pojmowaniu badań ewaluacyjnych w edukacji* [On the permanent need of breaking reductionism in understanding evaluation studies in education], [in:] B. Niemierko, M. K. Szmigel (eds.), *Regionalne i lokalne diagnozy edukacyjne*, PTDE, Kraków 2012.

[16] Krope P., *Ocena opisowa w pedagogice jako źródło nieporozumień* (Educational Commented report as a source of misunderstanding], [in:] B. Niemierko red., *Diagnostyka edukacyjna*, Wydawnictwo UG, Gdańsk 1994.

[17] Messick S., *Test validity and the ethics of assessment*, „American Psychologist", 1980, 35.

[18] Messick S., *Validity*, [in:] R. L. Linn red., *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989.
[19] Niemierko B., *Ogólnopolskie badania osiągnięć uczniów, nauczycieli i szkół. Synteza badań* [National study of student, teacher, and school achievement. Research synthesis], Centrum Doskonalenia Nauczycieli, Warszawa 1990.
[20] Niemierko B., *Pomiar wyników kształcenia* [The measurement of teaching outcomes], WSiP, Warszawa 1999.
[21] Niemierko B., *Ocenianie szkolne bez tajemnic* [Grading without mystery], WSiP, Warszawa 2002.
[22] Niemierko B., *Kształcenie szkolne. Podręcznik skutecznej dydaktyki*. [School instruction. Handbook of efficient didactics], WAiP, Warszawa 2007.
[23] Niemierko B., *Educational diagnostics for contemporary school systems. Measuring and assessing growth of student human capital. Part I: Main concepts and the scope*, „Colloquium", 2012, 1.
[24] Niemierko B., *Educational diagnostics for contemporary school systems. Measuring and assessing growth of student human capital. Part II: Methodology and rules*, „Colloquium", 2012, 2.
[25] Patton M. Q., *Creative evaluation*. Sage, London 1981.
[26] Popham W. J., *Educational evaluation*. Prentice-Hall, Englewood Cliffs 1975.
[27] Robson C., *Real world research. A resource for social scientists and practitioner – researchers*, Blackwell, Oxford 1993.
[28] Shepard L. A., *Classroom assessment*, [in:] Brennan (ed.) *Educational measurement. Fourth edition*, American Council on Education – Praeger, Westport 2006.
[29] Smith J. K., *Reconsidering reliability in classroom assessment and grading*. „Educational Measurement: Issues and Practice", 2003, 4.
[30] Suchman E. A., *Evaluative research: principles in public service and action programs*, Sage, New York 1967.
[31] Szyling G., *Nauczycielskie praktyki oceniania poza standardami* [Teachers' practices of assessment out of standards], Impuls, Kraków 2011.

[32] Willingham W. W., Pollack J. M., Lewis C., *Grades and test scores: Accounting for observed differences,* „Journal of Educational Measurement", 2002, 1.
[33] Ziemski S., *Problemy dobrej diagnozy* [Problems of a valid diagnosis], Wiedza Powszechna, Warszawa 1971.

DIAGNOSTYKA EDUKACYJNA WE WSPÓŁCZESNYM SYSTEMIE SZKOLNYM. POMIAR I OCENIANIE WZROSTU KAPITAŁU LUDZKIEGO UCZNIÓW. CZĘŚĆ III: WYBRANE WYNIKI

STRESZCZENIE

Praktyka edukacyjna obfituje w myślenie życzeniowe, nasycone przesadnym optymizmem. Na przykład w Polsce stosujemy bardzo niskie (30%) ilościowe normy wymagań w egzaminach zewnętrznych, a wierzymy, że ich osiągniecie oznacza opanowanie przedmiotu.

Pomyślne dla diagnostyki edukacyjnej jest to, że ocenianie wewnątrzszkolne obejmuje, obok poznawczych, emocjonalno-motywacyjne osiągnięcia uczniów. To czyni je dość dobrym predyktorem sukcesów tych uczniów w dalszym kształceniu i w pracy zawodowej.

Ocenianie przyrostu kapitału ludzkiego w kształceniu ogólnym zaczyna dopiero dochodzić do głosu na świecie i w Polsce, często w ukrytej, nie w pełni legalnej formie „drugiego układu wymagań egzaminacyjnych". Na niższych szczebla systemu edukacyjnego przybiera to zwykle postać oceny opisowej, w której nauczyciele ujmują osiągnięcia ucznia własnymi słowami, mocno wykraczając poza obszar wiadomości i umiejętności.

Zarówno pedagodzy, jak i uczniowie, są nastawieni krytycznie wobec sprawiedliwości obecnych praktyk oceniania wewnątrzszkolnego, co naraża trafność konsekwencyjną tego oceniania na uzasadnione zarzuty. Z kolei próby rejestrowania zjawisk pedagogicznych w drodze ogólnosystemowych badań o charakterze ewaluacyjnym mogą być skażone wpływem różnych czynników politycznych. Poszerzenie diagnozy

edukacyjnej, tak by objęła wszystkie aspekty rozwoju kapitału ludzkiego, mogłoby zaowocować działaniami o większej wartości ekonomicznej i etycznej.

Słowa kluczowe:
diagnostyka edukacyjna, myślenie życzeniowe, osiągnięcia uczniów, ocena opisowa, ocenianie szkolne, trafność konsekwencyjna, ewaluacja w edukacji.