

Bolesław Niemierko
Akademia Marynarki Wojennej w Gdyni

**EDUCATIONAL DIAGNOSTICS FOR
CONTEMPORARY SCHOOL SYSTEMS.
MEASURING AND ASSESSING GROWTH
OF STUDENT HUMAN CAPITAL
PART II: METHODOLOGY AND RULES**

ABSTRACT

Developmental paradigms, preferred in educational diagnostics to intervening ones, need vertical scaling, *i.e.* norms extended across a time passage or a learning cycle. Value added is perceived in education as a progress made by a student or a group of students in a period of time in a well-defined area of attitudes or skills. It may be evaluated in raw scores, percentiles, stanines, grade equivalents or logits (theta scale).

Grade equivalents (*GE*) are numbers of years and months of schooling to yield given achievement levels. They are easy to calculate but possibly misleading. A layman may abuse them and suggest allocating students into school grades according to their *GE* indexes what would inevitably destroy the educational system. Much more statistically advanced scaling is based on Item Response Theory (*IRT*) which is a probabilistic theory concerning the relation of an item score with the human property constituting a latent variable. We may apply *IRT* to mental test items, interview questions, behavioral categories, and even factual information obtained in document analyses.

The most influential context variable in educational diagnostics is socio-economic status of a student's family (*SES*) as indicated by the parents' education and vocational positions, their income level and social prestige. Both educational aspirations and examination scores are to large extent determined by *SES*. Two further unfavorable phenomena of school learning are intellectual helplessness which appears when a student's efforts to master the content of one or more school subjects proves totally unsuccessful, and learned one-sidedness that is a damage to a student's personality caused by his/her aspiration to be the best learner at school.

Most educational systems assume „equalizing educational chances for every child”,

„diminishing the gap between the best and the weakest” and making „no child left behind”. However, the politically fair ideas cannot stand psychological law of fan effect, that is an increase of achievement variance which comes with achievement growth in a population.

Keywords:

educational diagnostics, grade equivalent, Item Response Theory, socio-economic status, intellectual helplessness, learned one-sidedness, fan effect.

INFORMAL DIAGNOSTIC METHODOLOGY

In the previous part of the elaboration¹, four paradigms of educational diagnostics were marked out: informal intervening diagnostics, standardized intervening diagnostics, standardized intervening diagnostics, and standardized developmental diagnostics. Each paradigm has an original methodological armory but the distinction between informal and standardized approaches appears principal.

Informal diagnostic methods are based on *observation*. In most cases it is merely unstructured continuous watch. Its quality is strictly dependent upon the observers' professional competence. In most cases educational diagnostic observation takes shape of *participant observation* in which the observer plays significant role in triggering and controlling student activities what makes the method demanding and complex². The skill of looking at students' behaviors and thinking about the behaviors is crucial for educational leadership³. The main difficulty is speed of the classroom events which may be unexpected and challenging to a teacher. Could we expect her to deliberate upon a case and make exact notes in the situation of a war correspondent when every moment brings in „death risk” of making educational error and losing authority? Diagnostic skill of teachers grows slowly in the

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.

² K. Doktor, *O stosowalności obserwacji uczestniczącej* [On usability of participant observations], „Studia Socjologiczne”, 1961, 2.

³ T. L. Good, J. E. Brophy, *Looking in classroom*, Harper & Row, New York 1987, p. 49.

course of practice and educational observation skill of beginners may remain far behind their presentation skill⁴.

Another common method of informal diagnostics is *talk* with a student or a group of students on the topics of mutual interest. Interviews also require well prepared educators since students immediately „cross out” those educators who try to affect them neither knowing facts about the case nor the technicalities of verbal inquiry. As Thomas Gordon puts it:

Students are unusually sensitive about nonverbal communicates conveyed to them by teachers. They are able to read muscles tension, mouth tightness, face expression, and body movement. (...) It is almost impossible to hide the true emotions. (...) Finally, real emotions will always be disclosed⁵.

One of the most efficient means of diagnostic talk is *active listening*, paraphrasing student expressions without any comment⁶. Most student problems are emotional in nature. To penetrate them, one should (1) create a favorable situation, since „those who have real problems not always belong to the most talkative”⁷, (2) apply neutral *talk openers* („Would you like to talk to me about it?”), (3) avoid blockades like too hurried evaluation of events or entering into domains of value systems and culture differences.

Educational interview is based on mutual *trust* and to be successful a teacher must secure „high account in emotional bank of the classroom”⁸. So diagnostic talk is always double-connected to *the past*: by the topic of previous events and by the history of personal familiarity between the interlocutors.

Document analysis may be used as source of vital information on student biography and achievement. Documents include certificates, portfolios

⁴ M. Maciaszek, *Kształtowanie umiejętności dydaktycznych nauczyciela* [Shaping teacher's instructional skills], PWN, Warszawa 1965.

⁵ T. Gordon, *Teacher effectiveness training*, Gordon Training International, Solana Beach 1974. Polish translation: *Wychowanie bez porażek w szkole*, PAX, Warszawa 1997, p. 47.

⁶ T. Gordon, *Parent effectiveness training. The tested new way to raise responsible children*, Gordon Training International, Solana Beach 1970; E. Bochno, *Rozmowa jako metoda oddziaływania wychowawczego* [Talk as a method of educational action], Impuls, Kraków 2004; A. Suchańska, *Rozmowa i obserwacja w diagnozie psychologicznej*. [Talk and observation in psychological diagnosis], WAiP, Warszawa 2007.

⁷ A. Janowski, *Poznanie uczniów. Zdobywanie informacji w pracy wychowawczej*. [Recognizing students. Gaining information in upbringing work], WSiP, Warszawa 1975. New edition: *Fraszka Edukacyjna*, Warszawa 2002, p. 140.

⁸ S. R. Covey, *Synergic communication*, [in:] J. Stewart (ed.), *Bridges not walls. A book about interpersonal communication*, Mc-Graw-Hill, New York 1995. Polish translation: *Komunikacja synergiczna*, [w:] J. Stewart red., *Mosty zamiast murów. O komunikowaniu się między ludźmi*, PWN, Warszawa 2000, p. 67.

of student products⁹, various types of homework, synectic self-projections¹⁰, school official reports, personal diaries, blogs, letters, pictures, and various objects of regular use. Their interpretation in education is mainly informal since the structured methods of content analysis are too burdensome and time-consuming, yet certain questions should be asked before using them in diagnostic procedures:

1. When and where the document was created? Who was the real author of the document?

2. What was the purpose of the document: natural (internal need) or dictated (required) by authorities or specific situation?

3. What was the competence of the author (knowledge of facts, communication proficiency) in the field?

4. Are we in full readiness to interpret the document (sufficient familiarity with the authors, firsthand knowledge of context, analytical skill)?

5. Can we prove the facts stated in the document and the validity of our interpretations?

Inquiry (questionnaire) is a method of gaining information by asking questions and receiving answers in written form. It is usually applied to group problems in anonymous way in a classroom and is expected to provide students with self-portrait of the class in meaningful aspects of their life. Here are some examples of questions connected to the human and social capital of the youth¹¹:

1. (learning environment) *What are your regular duties at home?* (Are there any? Are you used to responsible work on your own? Are not you overloaded with home duties?)

2. (learning experiences) *Describe three events in your life which taught you something important.* (Were they school-based or independent? Individual or group-related?)

3. (peer group) *What one should do or possess to gain high position in your colleague group?* (Is it a matter of family social position or rather of the student's individual traits and effort? Do you want to lift your position? Do you know how to move it up?)

⁹ S. Valencia, *Portfolios: Panacea or Pandora box?*, [in:] F. L. Finch (ed.), *Educational performance*, Riverside, Chicago 1991.

¹⁰ W. J. Gordon, *Synectics*, Harper, New York 1961.

¹¹ A. Janowski, *Poznanwanie uczniów. Zdobywanie informacji w pracy wychowawczej*, [Recognizing students. Gaining information in upbringing work], WSiP, Warszawa 1975.

4. (plans for future) *Which of school subjects will likely come out the most suitable for your future life after your school education is finished?* (Does school education matter to the students? What kinds of knowledge and skills they regard as beneficial for their future careers? Do they self-regulate development of the knowledge and skills?)

The last of the key methods of informal diagnostics is *experiment*, more precisely – *pre-experiment (quasi-experimental design)*, without sampling, randomization, exact stimuli control, and multiple measurement, a kind of *action research*, in natural conditions¹². Pre-experiment is ubiquitous in non-scientific human actions, where outcomes are evaluated after every original activity. As Earl Babbie puts it:

We continuously experiment in our trials of working out general explanation of the world we live in. All our skills are acquired through experiments: eating, walking, speaking, riding bicycle, swimming. Experimenting, students learn how much they have to learn to be successful at the university. Through experiments professors learn how long they have to prepare for a good lecture¹³.

Most educational experiments are almost automatically successful by paying attention to a problem and by increasing effort to solve it, what is sometimes called *the Hawthorn effect*¹⁴. However, some of educational experiments are evidently unsuccessful. It was the case of Janusz Korczak's experiment on children's independent judiciary. The children courts appeared schools of lying, cheating, and cowardice rather than a method of improving moral education. Korczak concluded the experiment with a reflection upon the lesson he learned as an educational diagnostician:

One court case often gave me better child's characteristics than several months of meeting him. Sometimes one case gave me better characteristic of environment than free observation in a couple of months. As the court sec-

¹² D. T. Campbell, J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago 1966; C. Frankfort-Nachmias, D. Frankfort, *Research methods in social sciences*, Scientific American/St. Martin's College 1966. Polish translation: *Metody badawcze w naukach społecznych*, Zysk, Poznań 2001, p. 162-166.

¹³ E. Babbie, *The practice of social research. Ninth edition*, Wadsworth/Thomson Learning, Belmont 2001. Polish translation: *Badania społeczne w praktyce*, PWN, Warszawa 2004, p. 246.

¹⁴ F. I. Roethlisberger, W. T. Dickson, *Management and the worker*, Harvard University Press, Cambridge 1939.

retary I got to know the alphabet, improved myself, and became an expert in their problems¹⁵.

Most informal diagnostic methods: observation, talk, document analysis, and pre-experiment, are involved in low-stake, everyday *student achievement assessment*. Teachers use them for grading purposes in combined, mixed, and sometimes even „hodge-podge” ways¹⁶. Usually, *formative assessment*, providing „information about the learning process that teachers can use for instructional decisions and students can use for improving their own performance, which motivates students”¹⁷ prevails over *summative assessment*, oriented toward achievement certification. Needs for immediate feedback shorten the distance between informal diagnosing and teaching in education¹⁸.

STANDARDIZED DIAGNOSTIC METHODOLOGY

Standardized diagnostic methods are based on *measurement*, mainly of its *norm-referenced* type, where student score is interpreted by means of population score distribution. The *criterion-referenced* tests, in which a student score is interpreted according to educational standards, prove less usefulness in standardized diagnostics as they are too subjective in standard setting procedure and may tell us more about the standard developers than on the students themselves¹⁹.

There is a variety of sophisticated measurement tools to be used in educational standardized diagnostics: achievement tests, personality tests, questionnaires, rating scales, descriptive scales, coding sheets, content analytical schemes. Not only their constructors but also their users need thorough theoretical background and intensive practical training.

¹⁵ J. Korczak, *Jak kochać dziecko* (How to love a child), [in:] J. Korczak, *Wybór pism*, Tom III, Nasza Księgarnia, Warszawa 1921/1958, p. 382.

¹⁶ S. M. Brookhart, *Grading*, Pearson, Upper Saddle River 2004.

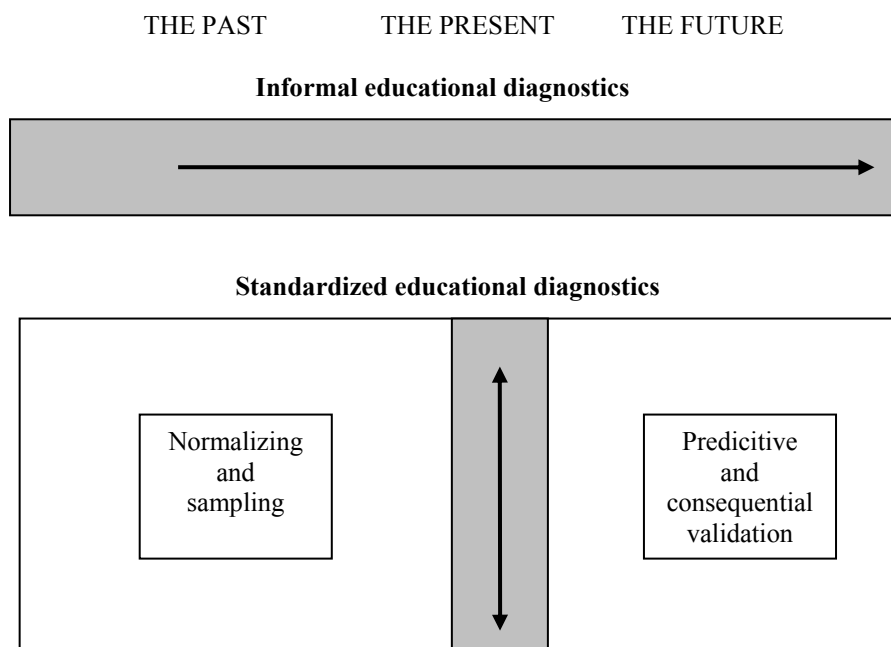
¹⁷ S. M. Brookhart, *Editorial*, „Educational Measurement: Issues and Practice”, *Special Issue: The Validity of Formative and Interim Assessment*, 2009, 1.

¹⁸ J. P. Leighton, M. J. Gierl (ed.), *Cognitive diagnostic assessment. Theory and applications*, Cambridge University Press, New York 2007.

¹⁹ B. Niemierko, *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania* [Criterion-referenced measurement in education. Theory and applications], PWN, Warszawa 1990; R. K. Hambleton, *The rise and fall of criterion-referenced measurement?*, „Educational Measurement: Issues and Practice”, 1994, 4.

Standardized educational diagnostic procedures are distinct from the informal ones in length and width of working contacts between the people involved. Figure 1 illustrates how profound is the difference.

Figure 1. Comparison between informal and standardized diagnostics



Informal diagnoses concern individuals and small groups of students in their past, present, and future situations where regular contacts with educators are daily occurrence. Such diagnoses are *longitudinal*, they link information on consecutive stages of youth development, what is symbolized by one-point arrow in Figure 1. This feature increases assessment reliability²⁰ but some flaws appear as well: context-related opinions, halo-effects, rater instabilities. These flaws are unavoidable since educational action research combines cognition and impact on student learning in one inseparable unit of teacher activity.

Standardized diagnosis operates in current reality and both the past and the future have only methodological importance. It is *cross-sectional*,

²⁰ J. K. Smith, *Reconsidering reliability in classroom assessment and grading*, „Educational Measurement: Issues and Practice”, 2003, 4.

transverse and usually large-scale procedure what is symbolized by two-point arrow in Figure 1. Population norms enable us to interpret student scores on account of learning abilities and learning outcomes in precisely defined, significantly distinct²¹ and appropriately functional areas. Thanks to standardization processes, objectivity, reliability, and validity of conclusions are in control.

Most standardized procedures are supplemented by „soft” informal observation of behaviors and many informal methods have certain „hard” elements, like written directions, uniform tasks, pseudo-norms²². The boundary between two approaches is movable according to (1) the type of problem, (2) time at disposal, (3) population size, (4) availability of measurement tools, and (5) the diagnostician’s competence. Figure 3 presents the method selecting area as an interval on the diagnosis certainty line. It follows James Popham’s²³ illustration of specifying test outline.

Figure 2. Optimizing methods of educational diagnosis

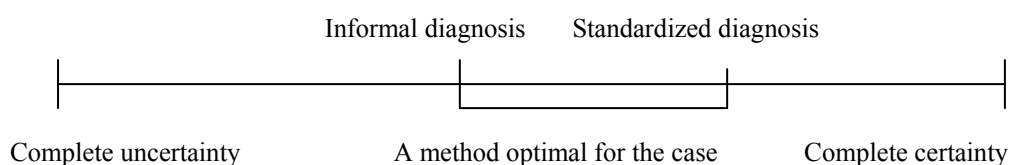


Figure 2 shows that:

1. *Certainty of diagnosis*, as a part of its validity²⁴, is a continuous variable.
2. Every kind of educational diagnosis considerably reduces uncertainty of learning processes.
3. Informal diagnosis brings substantial reduction of doubts.
4. Standardization increases certainty of diagnosis.
5. Complete certainty of educational diagnosis is unavailable.

²¹ S. Sinharay, *How often do subscores have added value? Results from operational and simulated data*, „Journal of Educational Measurement”, 2010, 2, p.150-174.

²² B. Niemierko, *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania* [Criterion-referenced measurement in education. Theory and applications], PWN, Warszawa 1990, p. 275-279.

²³ W. J. Popham, *Domain specific strategies*, [in:] R. A. Berk (ed.), *Criterion-referenced measurement. The state of art*, John Hopkins University Press, Baltimore 1980.

²⁴ S. Messick, [in:] R. L. Linn (ed.), *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989.

VERTICAL SCALING

Developmental paradigms of educational diagnostics need measurement scales capable of displaying a student's progress in major domains of human capital. While scores obtained within an age or school grade cohort are horizontal in range, the scores that span several successive cohorts enable vertical interpretation. *Vertical scaling* results in norms extended across a time passage or a learning cycle and make possible estimation of relevant value added.

Value added, the term borrowed from economy, is considered in education as a progress made by a student or a group of students in a period of time in a well-defined area of attitudes or skills. It may be evaluated in raw scores, percentiles, stanines, grade equivalents, logits or in other units of ordinal or interval scales. Grade equivalents, as the easiest to interpret, and logits, as the most statistically advanced, will be discussed here.

Grade equivalents (GE) are numbers of years and months of schooling to yield given achievement levels. They may be calculated in test standardization procedures as vertical scaling which links test scores with successive grades.

Grade equivalents constitute an additive scale in terms of increase in student ability in specified domain of school achievement. The scale may be used to compare the student's yearly progress and, thanks to considerable reduction of measurement error when groups rather than individuals are tested, the monthly school progress in selected subjects or abilities.

When developmental goals of education are emphasized, student achievement may be treated as cumulative what means that test scores can be used to make comparisons over time and to measure student growth as they move on learning continuum. Essential academic skills, such as reading, writing, mathematical reasoning, understanding science, and understanding social environment, are suitable for his kind of cumulative interpretation²⁵. These skills may be measured with parallel tests over school grades, so vertical scales, based on anchoring successive grades by common items, testlets or whole tests, may be constructed.

²⁵ M. J. Kolen, *Scaling and norming*, [in:] R. L. Brennan (ed.), *Educational measurement. Fourth edition*, American Council on Education – Praeger, Westport 2006; J. Tomkowicz, *Development and use of vertical scales to measure student achievement*, [in:] B. Niemierko i M. K. Szmigiel (ed.), *Uczenie się i egzamin w oczach nauczycieli*, PTDE, Kraków 2008.

When tests are vertically standardized, the vertical scale they constitute is straightforward to interpret at the whole range of $K - 12$, what means „from kindergarten to high school graduation”. The author’s experience with junior high school (gymnasium) students in Poland revealed that standard deviation of individual *GEs* at this level of education was about 1.5 year and the value added for student groups of similar (low, medium, or high) ability in this three-year school ranged from 2.1 to 3.9, and for entire, intact school clusters ranged from 2.4 to 3,5²⁶. Since the progress indexes accumulate across the school levels the *GE* for the junior high-school graduate averages extended from 7.5 (1.5 year behind the national norm) to 11,1 (2.1 above the norm). Differences of this size must make us anxious about the existence of democratic educational system in our country²⁷.

A layman would likely abuse grade equivalents and suggest allocating students into school grades according to their *GE* indexes. Naïve, too far reaching conclusions of this kind threaten school systems with destruction and this is why developmental paradigms and scales require fully conscious, well-prepared users²⁸.

On the other hand, fear of classifying humans in a cool, context-free, ruthless way would prevent educators from any application of educational diagnostics and make human capital a purely metaphoric term. A chance of adjusting conditions and treatments to individual differences in abilities and interests would be lost. As a guiding rule, *universal valuing* of people as „better” or „worse” should be banned from educational diagnostics but it does not limit inquiries about their *strengths and weaknesses* in order to support their learning.

As human capital dimensions, measurement scales should cover the whole length of variability, from appearance of a disposition with newborns to full mastery of appropriate actions at the expert level. Such long scales were built as early as in the eighties²⁹ to measure competencies along the „novice – expert” continuum. Their significance increased in the next dec-

²⁶ B. Niemierko, *Grade equivalent as a measure of value added for student and school achievement*, „Polish Journal of Social Science”, 2008, p. 43-58.

²⁷ R. Dolata, *Szkola – segregacje – nierówności* [School – segregations – inequalities], Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2008.

²⁸ H. B. Lyman, *Test scores and what they mean*, Prentice-Hall, Englewood Cliffs 1963; A. J. Nitko, *Educational tests and measurements. An introduction*, Harcourt, New York 1983.

²⁹ J. Larkin, J. McDermott, D. P. Simon, H. A. Simon, *Expert and novice performance in solving physics problems*, „Science”, 1980, p. 1335-1342.

ades when Item Response Theory appeared to prevail over the classical test theory³⁰. Although implementations of the new measurement theory into the Polish educational system started on just now³¹ (Szaleniec, 2010), there is no doubt about its future usefulness to educational diagnostics.

NEW APPLICATIONS OF IRT-SCALING

Item Response Theory (IRT) is a probabilistic theory concerning the relation of an item score with the human property measured by a set of items. This property constitutes a *latent variable*, carefully defined at the stage of test outlining but unobservable, manifested itself only by the specific pattern of item scores. To make use of the pattern we have to accept an assumption that the scores are *locally independent* what means that the latent variable and the measurement error are the only sources of item score variability. The accuracy of matching the model to score distributions should be empirically proved³².

All methods of educational diagnosis, from unstructured observation to the most sophisticated tests, may be used to select highly discriminative, coherent items to construct sufficiently homogeneous measurement scales. In *theta* (θ) scaling of a latent variable an average person has probability $p = 0.50$ of passing the item in the middle of the scale, a lower probability when he/she is less able or the item is more difficult, and a higher probability on the reverse of that.

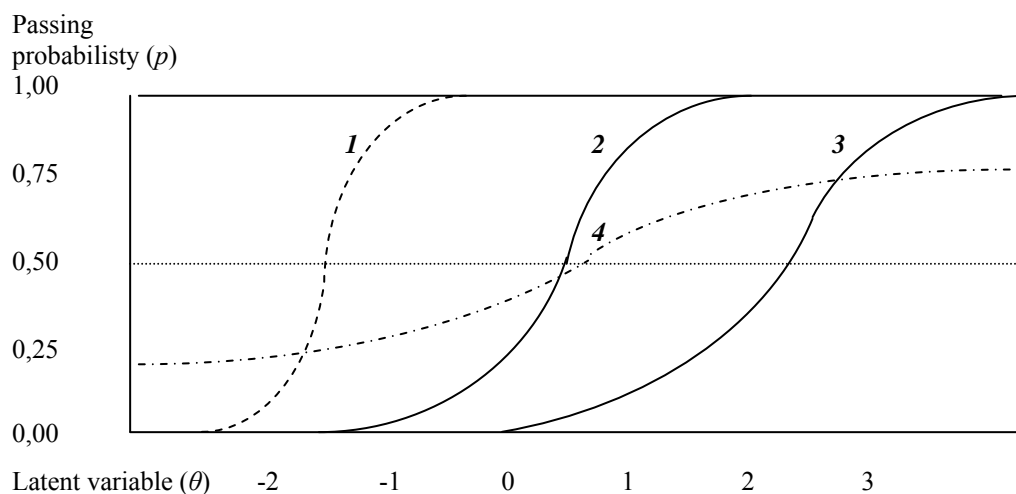
In *two-parameter logistic IRT measurement model (2PL)* the θ value of the latent variable is a function of a person ability and an item difficulty. The *item characteristic function*, $f(\theta)$, of four exemplary items in Figure 3 will show us how it works.

³⁰ R. K. Hambleton, *Comparison of Classical Test Theory and Item Response Theory and their applications to test development*, „Educational Measurement: Issues and Practice”, 1993, 3.

³¹ H. Szaleniec (ed.), *Teoria wyniku zadania, IRT. Zastosowania w polskim systemie egzaminów zewnętrznych* [Item Response Theory. Applications to the Polish system of external examinations], Centralna Komisja Egzaminacyjna, Warszawa 2010.

³² R. K. Hambleton, *Principles and selected applications of item response theory*, [in:] R. L. Linn (ed.), *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989, p. 172-182.

Figure 3. Item characteristic functions



The two *item parameters*, *i.e.* its determining and constant features, are following:

1. *Discrimination*, a , is a slope of the curve at the point of $p = 0.50$. It may assume values from minus infinity to plus infinity but in practice appears between $a = 0.5$ and $a = 2.5$ ³³. Four curves in Figure 3 indicate the following approximate values: $a_1 = 2.5$ (excellent discrimination); $a_2 = 2.0$ (good discrimination); $a_3 = 1.5$ (medium discrimination); $a_4 = 0.5$ (weak discrimination).

2. *Difficulty*, b , is a position of the point of $p = 0.50$ on the theta scale of latent variable. Also this parameter assumes values from minus infinity to plus infinity but in practice the interval $\{-3.00; +3.00\}$ suffices for item analysis and a measurement tool construction. Difficulty of items presented in Figure 3 is the following: $b_1 = -2.0$ (very easy item), $b_2 = 0.0$ (average difficulty), $b_3 = 2.0$ (very difficult item), and $b_4 = 0.0$ (average difficulty).

In most cases *IRT* is applied to mental test items but it is also suitable for interview questions, behavioral categories, and even for factual information obtained in document analyses. Every index that consistently discriminates students and job candidates at the anticipated level of proficiency adds to the information gathered in adequate interval of a theta scale. Thetas may

³³ E. Hornowska, *Testy psychologiczne. Teoria i praktyka*, [Psychological tests. Theory and practice], Scholar, Warszawa 2001, p. 198.

Observation research proves that *SES* differentiates types of students into two major groups³⁷:

1. Supposedly from the high-*SES* families: (1) *success students*, task-oriented and academically successful, (2) *dependent students*, seeking such teacher help and support as they receive at their own homes, and (3) *phantom students*, working independently at about average level, nearly invisible in their class work.

2. Supposedly from low-*SES* families descendants: (1) *social students*, who have the ability to achieve but prefer friendships and peer activities, and (2) *alienated students*, reluctant learners, who reject „everything for which school stands”, withdrawn and often hostile towards teachers³⁸.

Family influences the children’s level of *educational aspiration* which may be operationally defined as the standard of education they expect to achieve. It will be illustrated by a research conducted in all lower secondary schools (gymnasia) in one region of Poland³⁹. *SES* was divided into 5 groups: low (16%), lower middle (20%), middle (44%), higher middle (10%), and high (10%). Educational levels were specified as vocational basic, vocational high school, college, master degree, more than master degree. The research outcomes are presented in Table 1.

Table 1. Socio-economic status and educational aspirations (in percentages)

Children’s educational aspiration	Parents’ socio-economic position					Total
	low middle	lower middle high	middle	higher		
Vocational basic	2	1	0	-	-	1
Vocational high school	46	36	21	6	2	24
College	15	14	14	9	4	13
Master degree	29	35	45	59	59	43
More than master degree	8	14	20	26	35	19
Total	100	100	100	100	100	100

³⁷ T. L. Good, C. N. Power, *Designing successful classroom environments for different types of students*, „Journal of Curriculum Studies”, 1976, 8, p. 45-60.

³⁸ A. J. Nitko, *Educational tests and measurements. An introduction*, Harcourt, New York 1983, p. 277.

³⁹ J. Domalewski, P. Mikiewicz, *Młodzież w zreformowanym systemie szkolnym* [Youth in the reformed school system], IRWiR PAN, Toruń 2004.

There are two main conclusions of Table 1:

1. Aspirations of gymnasium grade 1 students are much higher than their parents' education level. As many as 60% of them intend for master and above-master education in stationary or extra-mural procedure. For some of them it may be a sort of fantasy but generally reflects a positive trend.

2. Aspirations are moderately correlated with socio-economic status (Pearson $C = 0.36$, $n = 4069$). Low-SES group prefers vocational education while the rest of the student population aspire in the majority to the master degree. The correlation may be interpreted as a causal relationship since it ordinarily happens that aspirations are *socially inherited*, derived from ancestors but never from descendants, especially with teenagers.

We do not have information on the later academic and vocational careers of the students who were inquired about aspirations but we do know their final examination scores. The scores are presented on the standard five-point scale in Table 2.

Table 2. Socio-economic status and gymnasium finals score (in percentages)

Examination score	Parents' socio-economic position					Total
	low	lower middle	middle	higher middle	high	
Low	15	10	4	1	-	6
Lower middle	39	34	22	11	2	24
Middle	35	39	44	31	22	39
Higher middle	10	14	24	43	52	24
High	1	3	6	14	24	7
Total	100	100	100	100	100	100

We can see in Table 2 that examination scores also depend on the family SES and the correlation is even slightly higher ($C = 0.43$). The most characteristic quantities may be found at the extremes of the main diagonal of Table 4: the highest percentage of low scores belongs to low-status families and the highest percentage of high scores comes from high-status families. However, it happens that in low-SES families children accomplish above-middle score and, quite exceptionally, in high-SES families children fail to achieve at least middle score.

The data suggest that students' affective (aspirations) and cognitive (knowledge and skills) achievements are equally determined by their family SES. The two aspects of individual development are significantly intercorre-

lated (Domalewski and Mikiewicz report $C = 0.47$) and the statement of their mutual reinforcement stands to reason.

LEARNED HELPLESSNESS AND LEARNED ONE-SIDEDNESS

Most school systems strive for greater efficiency in cognitive domain what may generate students' emotional disturbance and may harm their learning. We will explore two kinds of the unwelcome results of the contemporary intensive education: learned helplessness and learned one-sidedness, two opposed effects of too hard pressure on academic success.

Intellectual helplessness appears when a student's efforts to master the content of one or more school subjects proved in vain. Helplessness is *learned* because feeling that the subject cannot be acquired results from a series of failures in learning and examinations. It is experienced by the students who originally were well motivated to learn but proved not equal to the demands, acknowledged that no effort would pay, and „behave as somebody who plunged into a deep well”⁴⁰. They lose physical energy and hope for the future, feel unhappy and depressed, sometimes under a mask of complacency or rebelliousness.

Learned helplessness was first identified with laboratory experiments on animals⁴¹, then studied on humans by means of questioning⁴². The typical items of attitude scales of intellectual helplessness at school learning are following⁴³:

1. *Emotional deficit*: I am ashamed of not understanding a lesson.
I am frightened of making fun of myself.
2. *Motivational deficit*: I feel I waste time at school.
Our lessons are deadly boring.
I cannot wait until the end of the lesson.

⁴⁰ G. Mietzel, *Pädagogische Psychologie des Lernens and Lehrens*, [Educational psychology of learning and teaching], Hogrefe-Verlag, Göttingen 2001, p. 371.

⁴¹ M. Seligman, *Helplessness: On depression, development, and death*, Freeman, San Francisco 1975.

⁴² G. Sędek, *Bezradność intelektualna w szkole* (Intellectual helplessness at school), Instytut Psychologii PAN, Warszawa 1995.

⁴³ B. Ciżkowicz, *Wyuczona bezradność młodzieży* [Learned helplessness of youth], Uniwersytet Kazimierza Wielkiego, Bydgoszcz 2010.

3. *Cognitive deficit*: I do not know the answer to the teacher questions.

I do not understand what the teacher explains.

The answers („never” – „rarely” – „sometimes” – „often” – „always”) to the above quoted and similar questions show that:

1. The percentage of students suffering from heavy intellectual helplessness increases from about 5% in elementary school to above 15% in higher education school.

2. Mother tongue in elementary school and mathematics in high school are the main sources of intellectual helplessness.

3. Boys are more susceptible to helplessness in elementary school and girls are more depression-inclined in the higher levels of education.

Learned one-sidedness (workaholic person) is the damage to a student's personality caused by his/her aspiration to be the best learner at school and the best employee at work. The damage was analyzed by Dorota Turska who studied 150 best students of lower high-school (gymnasium) and high-school (lyceum) in one of the regions in Poland. She distinguished two domains of differences between the best and the average students: emotional/social and cognitive/creative.

Table 3 presents the outcomes of questioning students about the *aims in life*, that is their main concerns in emotional perspective⁴⁴.

Table 3. The multitude of aims in life declared by the best and average students

The number of aims in life	The best students		Average students	
	gymnasium	liceum	gymnasium	liceum
One	51%	54%	24%	24%
Two or more	49%	46%	76%	76%

The best students appeared „concentrated upon school activity” which for the half of them was the only target of their efforts. „For tomorrow, next week, next month, next year, the nearest ten years, and the whole life”,

⁴⁴ D. Turska, *Skuteczność ucznia. Od czego zależy udana realizacja wymogów edukacyjnych* [Student efficacy. What are the prerequisites of successful realizing educational demands], UMCS, Lublin 2006.

as they were asked about, they consistently want to remain the best⁴⁵. Their self-esteem is higher than in average students but unsafe, requiring reinforcements (praises, rewards, admiration). They are success-dependant with low tolerance of failure and they seem to be less prepared to take troubles and to exceed their superiors' anticipation in further education and job.

The number of objects aimed at is considerably greater in average students. They more often appreciate family life, friends, social work, recreation and travel, adventures. They see effort of learning as a means rather than as their life design.

The cognitive perspective of the best students is also disappointing. Although they considerably surpass the average students in IQ measures (Raven), creativity tests bring diametrically opposed results. Table 4 presents the comparison⁴⁶.

Table 4. Intelligence and creativity in groups of the bests and average students

Trait	The best students		Average students	
	gymnasium	liceum	gymnasium	liceum
Intelligence	47,1	50,2	44,9	48,4
Creativity	20,1	25,4	22,5	26,7

Both traits grow from gymnasium to lyceum as a result of selection processes but the difference between them is invariable:

1. The best students are more capable of solving abstract, logical problems and of learning well-arranged, highly theoretical knowledge. Their chances for job success besides academic professions may be overestimated.

2. The average students are more creative in real-life, practical situations. Their predominant intelligence is often kinesthetic, spatial, musical, natural, interpersonal or intrapersonal rather than logical or verbal⁴⁷. Their successes are smaller at school and greater after graduation.

⁴⁵ Herein, p. 173.

⁴⁶ Herein, p. 83-84.

⁴⁷ H. Gardner, *Frames of mind*, Basic Books, New York 1993.

FAN EFFECT AND ATTEMPTS TO OVERCOME IT

Fan effect is an increase in achievement variance which comes with achievement growth in a population. Development rate positively correlates with the level of achievement, so generally it is higher for better students and lower for weaker students.

Most educational reforms affirm „equalizing educational chance for every child”, „diminishing the gap between the best and the weakest”, and making „no child left behind”⁴⁸. However, these politically fair ideas are unmercifully crushed in touch with psychological laws which relate learning rate to cognitive prerequisites⁴⁹. Time needed for learning a new skill at the higher level of competence is shorter than at the lower level, and consequently the better students are progressing faster with whatever teaching method and there is no significant „trait-treatment interaction”⁵⁰.

The Benjamin Bloom’s *mastery learning* theory was a trial of stopping and even reversing fan effect. He assumed that „most students become very similar with regard to learning ability, rate of learning, and motivation for further learning, when provided with favorable learning conditions”⁵¹ but this statement was never empirically proved. Only „the method of Robin Hood – to take time from the cognitively rich and to give it to the cognitively poor” could result in diminishing educational fan effect⁵².

To illustrate fan effect the national norms of a reading comprehension test⁵³ are presented in graphical form in Figure 5.

⁴⁸ *No Child Left Behind*, Pub Law Act No. 107-110, USA, 2002.

⁴⁹ See: 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, Fig. 1.

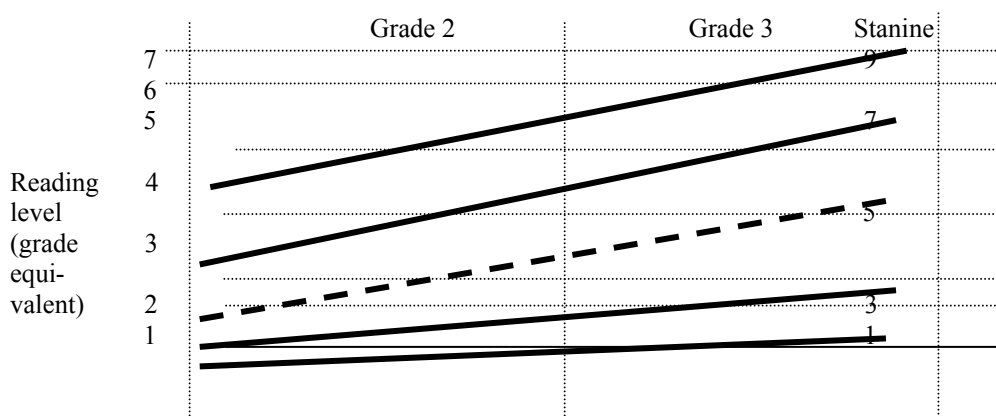
⁵⁰ L. J. Cronbach, R. E. Snow, *Aptitude and instructional methods. A handbook for research on interactions*, Irvington, New York 1977.

⁵¹ B. S. Bloom, *Human characteristics and school learning*, McGraw-Hill, New York 1976, p. X.

⁵² M. Arlin, *Time, equality, and mastery learning*, „Review of Educational Research”, 1985, p. 65-86.

⁵³ T. L. Kelley, R. Madden, E. F. Gardner, H. C. Rudman, *Stanford Achievement Test. Primary II Battery*, Harcourt, New York 1964.

Figure 5. Fan effect on a reading comprehension test



Thick lines in Figure 5 indicate mean scores on the reading comprehension test converted to *grade equivalents*, i.e. to the number of years and months of schooling necessary to make given test score the population mean. The broken line in the middle of the figure (stanine 5) displays the growth of exactly two years, from the beginning of Grade 2 to the end of Grade 3. For two years (twenty months of learning) the weakest students (stanine 1) progressed only five months in their reading level and the best students (stanine 9) progressed twenty five months, five times more than the weakest.

Fan effect of learning is of great importance in developing human and social capital by means of education. It concerns all domains of human capital: motivational, moral, experiential, and practical, and the whole variety of individual strengths and abilities. Teachers can modify the standard instructional procedures and adapt them to different needs and potentials of particular students and groups of students. A student's *ability profile*, an individual constellation of capacities, is equally valued because each student „has many ways to succeed in the educational system, and each is given credentials equally recognized by society”⁵⁴. Though yet not commonly applied, this approach has clear advantage over one-sided academic emphasis of nowadays teaching routine.

Academically nonproficient students and cognitively inefficient schools are under thorough consideration in the American nationwide project *No Child Left Behind*, the most powerful of great many attempts to diminish

⁵⁴ A. J. Nitko, *Educational tests and measurements. An introduction*, Harcourt, New York 1983, p. 278.

the gap between fast developing and slow developing subjects of educational programs. Yet closing the gap in twelve years appeared unrealistic and in the fifth year of *NCLB* execution a milder version, requiring that students are only „on track to proficiency”, was introduced⁵⁵. However, the remedy did not work either. The students who began the third grade at the cutpoint between partially proficient and unsatisfactory level had only one sixth chance of achieving proficiency and one third chance of maintaining at least partial proficiency by the tenth grade of their education⁵⁶. The expectation chart illustrating this regularity reminds Pediatric Growth Charts used to monitor height and weight of children relative to national norms⁵⁷. It is hardly possible that a slow-growing student will move up to the category of high-grow rates. Betebenner complains:

Today, achievement mandates are stipulated based upon the moral imperative of high standards for all children with little concern regarding the likelihood the students reaching these goals. Given current progress of students, it is unlikely that sustained levels of growth necessary to reach these standards will occur soon⁵⁸.

Some schools force students into enormously intensive learning to climb up the academic achievement ladder at the expense of their emotional balance. Table 5 presents student attitudes toward learning expressed in Likert five-point scale in two out of eleven rural junior-high schools in Gdańsk region in Poland⁵⁹. Gymnasium G3 was the weakest in the group with grade equivalent (GE) = 5.1 on the entrance examination (beginning of Grade 1), GE = 7.5 on the final examination (end of Grade 3) and achievement value added of 2.4 years what deepened its outsider position. Gymnasium G8 had the 7th position in the group (GE = 5,6) on the entrance examination and the 3rd position on the finals what resulted in achievement value added of 3.5 years, the best in the group.

⁵⁵ M. Spellings, *Secretary announces growth model pilot* (press release), Department of Education, Washington D. C., 2005: U.S.

⁵⁶ D. Betebenner, *Norm – and criterion-referenced student growth*, „Educational Measurement: Issues and Practice”, 2009, 4, p. 49.

⁵⁷ E. Wiley, *Editorial*, „Educational Measurement: Issues and Practice”, 2009, 4.

⁵⁸ *Norm – and criterion – referenced student growth*, p. 50.

⁵⁹ T. Kutajczyk, B. Przychodzeń, *Czynniki skuteczności kształcenia ogólnego w gimnazjach wiejskich* [The efficiency factors of general education in rural gymnasia], OKE, Gdańsk 2008.

Table 5. Indexes of student attitudes toward school environment in two rural schools

Inquiry item	Gymnasium G3	Gymnasium G8
<i>I like my homeroom teacher</i>	4.8	3.5
<i>We get together with teachers to plan and prepare school events</i>	4.4	3.2
<i>Our parents take part in important classroom events</i>	3.8	2.9
<i>I like attending school</i>	4.1	3.4
<i>I feel safe in my school</i>	4.7	4.2
<i>The social climate in our school the is right</i>	4.4	3.5
<i>Relations between teachers and students are good</i>	4.5	3.4
<i>Our teachers keep dropping hints about efficient learning</i>	4.6	3.1
<i>Our teachers are interested how I cope with learning demands</i>	4.3	3.2
<i>We apply the acquired knowledge into practice</i>	4.5	3.6
Total	4.4	3.4

As we can see in Table 5 all the attitudinal items give the slow academically developing G3 an advantage over the fast academically developing G8. It implies that moral costs of overcoming the fan effect destiny should always be taken into consideration.

BIBLIOGRAFIA

- [1] Arlin M., *Time, equality, and mastery learning*, „Review of Educational Research”, 1985.
- [2] Babbie E., *The practice of social research. Ninth edition*, Wadsworth/Thomson Learning, Belmont 2001, Polish translation: *Badania społeczne w praktyce*, PWN, Warszawa 2004.
- [3] Betebenner D., *Norm – and criterion – referenced student growth*, „Educational Measurement: Issues and Practice”, 2009, 4.
- [4] Bloom B. S., *Human characteristics and school learning*, McGraw-Hill, New York 1976.
- [5] Bochno E., *Rozmowa jako metoda oddziaływania wychowawczego* [Talk as a metod of educational action], Impuls, Kraków 2004.
- [6] Brookhart S. M., *Grading*, Pearson, Upper Saddle River 2004.

- [7] Brookhart S. M., *Editorial*, „Educational Measurement: Issues and Practice”, *Special Issue: The Validity of Formative and Interim Assessment*, 2009, 1.
- [8] Campbell D. T., Stanley J. C., *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago 1966.
- [9] Cizkowicz B., *Wyuczona bezradność młodzieży* [Learned helplessness of youth], Uniwersytet Kazimierza Wielkiego, Bydgoszcz 2010.
- [10] Covey S. R., (1995) Synergic communication, [in:] J. Stewart (ed.) *Bridges not walls. A book about interpersonal communication*, McGraw-Hill, New York 1995, Polish translation: 2000, *Komunikacja synergiczna*, [in:] J. Stewart red., *Mosty zamiast murów. O komunikowaniu się między ludźmi*, PWN, Warszawa 2000.
- [11] Cronbach L. J., Snow R. E., *Aptitude and instructional methods. A handbook for research on interactions*, Irvington, New York 1977.
- [12] Doktor K., *O stosowalności obserwacji uczestniczącej* [On usability of participant observations], „*Studia Socjologiczne*”, 1961, 2.
- [13] Dolata R., *Szkoła – segregacje – nierówności* [School – segregations – inequalities], Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2008.
- [14] Domalewski J., Mikiewicz P., *Młodzież w zreformowanym systemie szkolnym* [Youth in the reformed school system], IRWiR PAN, Toruń 2004.
- [15] Frankfort-Nachmias C., Frankfort D., (1996) *Research methods in social sciences*, Scientific American/St. Martin’s College, Polish translation: 2001, *Metody badawcze w naukach społecznych*, Poznań: Zysk.
- [16] Good T. L., Power C. N., *Designing successful classroom environments for different types of students*, „*Journal of Curriculum Studies*”, 1976, 8.
- [17] Good T. L., Brophy J. E., *Looking in classroom*, Harper & Row, New York 1987.
- [18] Gordon T., *Parent effectiveness training. The tested new way to raise responsible children*, Gordon Training International, Solana Beach 1970.
- [19] Gordon T., *Teacher effectiveness training*, Gordon Training International, Solana Beach 1974, Polish translation: *Wychowanie bez porażek w szkole*, PAX, Warszawa 1997.
- [20] Gordon W. J., *Synectics*, Harper, New York 1961.

- [21] Hambleton R. K., *Principles and selected applications of item response theory*, [in:] R. L. Linn (ed.), *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989.
- [22] Hambleton R. K., *Comparison of Classical Test Theory and Item Response Theory and their applications to test development*, „Educational Measurement: Issues and Practice”, 1993, 3.
- [23] Hambleton R. K., *The rise and fall of criterion-referenced measurement?* „Educational Measurement: Issues and Practice”, 1994, 4.
- [24] Hornowska E., *Testy psychologiczne. Teoria i praktyka*. [Psychological tests. Theory and practice], Scholar, Warszawa 2001.
- [25] Janowski A., *Poznawanie uczniów. Zdobywanie informacji w pracy wychowawczej*, [Recognizing students. Gaining information in upbringing work], WSiP, Warszawa 1975, New edition: Fraszka Edukacyjna, Warszawa 2002.
- [26] Kelley T. L., Madden R., Gardner E. F., Rudman H. C., *Stanford Achievement Test. Primary II Battery*, Harcourt, New York 1964.
- [27] Kolen M. J., *Scaling and norming*, [in:] R. L. Brennan (ed.), *Educational measurement. Fourth edition*, American Council on Education – Praeger, Westport 2006.
- [28] Konarzewski K., *Uczeń* [Student], [w:] red. K. Konarzewski, *Sztuka nauczania. Szkoła* [The art of teaching. School], PWN, Warszawa 2004.
- [29] Korczak J., *Jak kochać dziecko* (How to love a child), [w:] J. Korczak *Wybór pism, Tom III*, Nasza Księgarnia, Warszawa 1921/1958.
- [30] Kutajczyk T., Przychodzeń B., *Czynniki skuteczności kształcenia ogólnego w gimnazjach wiejskich* [The factors of efficiency of general education in rural gymnasia], Okręgowa Komisja Egzaminacyjna, Gdańsk 2008.
- [31] Larkin J., McDermott J., Simon D. P., Simon H. A., *Expert and novice performance in solving physics problems*, „Science”, 1980.
- [32] Leighton J. P., Gierl M. J. (eds.), *Cognitive diagnostic assessment. Theory and applications*, Cambridge University Press, New York 2007.
- [33] Lyman H. B., *Test scores and what they mean*, Prentice-Hall, Englewood Cliffs 1963.
- [34] Maciaszek M., *Kształtowanie umiejętności dydaktycznych nauczyciela* [Shaping teacher's instructional skills], PWN, Warszawa 1965.

- [35] Messick S., *Validity*, [in:] R. L. Linn (ed.), *Educational measurement. Third edition*, American Council on Education – Macmillan, New York 1989.
- [36] Mietzel G., *Pädagogische Psychologie des Lernens and Lehrens*. [Educational psychology of learning and teaching], Hogrefe-Verlag, Göttingen 2001.
- [37] Niemierko B., *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania* [Criterion-referenced measurement in education. Theory and applications], PWN, Warszawa 1990.
- [38] Niemierko B., *Grade equivalent as a measure of value added for student and school achievement*, „Polish Journal of Social Science”, 2008.
- [39] 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.
- [40] Nitko A. J., *Educational tests and measurements. An introduction*, Harcourt, New York 1983.
- [41] *No Child Left Behind*, Pub Law Act No. 107-110, USA, 2002.
- [42] Popham W. J., *Domain specific strategies*, [in:] R. A. Berk (ed.), *Criterion-referenced measurement. The state of art*, John Hopkins University Press, Baltimore 1980.
- [43] Roethlisberger F. I., Dickson W. T., *Management and the worker*, Harvard University Press, Cambridge 1939.
- [44] Seligman M., *Helplessness: On depression, development, and death*, Freeman, San Francisco 1975.
- [45] Sędek G., *Bezradność intelektualna w szkole* (Intellectual helplessness at school), Instytut Psychologii PAN, Warszawa 1995.
- [46] Sinharay S., *How often do subscores have added value? Results from operational and simulated data*, „Journal of Educational Measurement”, 2010, 2.
- [47] Smith J. K., *Reconsidering reliability in classroom assessment and grading*, „Educational Measurement: Issues and Practice”, 2003, 4.
- [48] Spellings M., *Secretary announces growth model pilot* (press release), Department of Education, Washington D. C., 2005: U. S.
- [49] Suchańska A., *Rozmowa i obserwacja w diagnozie psychologicznej*. [Talk and observation in psychological diagnosis], WAiP, Warszawa 2007.
- [50] Szaleniec H. (ed.), *Teoria wyniku zadania, IRT. Zastosowania w polskim systemie egzaminów zewnętrznych* [Item Response Theory. Ap-

- plications to the Polish system of external examinations], Centralna Komisja Egzaminacyjna, Warszawa 2010.
- [51] Tomkowicz J., Development and use of vertical scales to measure student achievement, [in:] B. Niemierko, M. K. Szmigel (eds.), *Uczenie się i egzamin w oczach nauczycieli*, PTDE, Kraków 2008.
- [52] Turska D., *Skuteczność ucznia. Od czego zależy udana realizacja wymogów Edukacyjnych* [Student efficacy. What are the prerequisites of successful realizing educational demands], UMCS, Lublin 2006.
- [53] Valencia S., *Portfolios: Panacea or Pandora box?*, [in:] F. L. Finch (ed.), *Educational performance*, Riverside, Chicago 1991.
- [54] Wiley E., *Editorial*, „Educational Measurement: Issues and Practice”, 2009, 4.

DIAGNOSTYKA EDUKACYJNA WE WSPÓLCZESNYM SYSTEMIE SZKOLNYM. POMIAR I OCENIANIE WZROSTU KAPITAŁU LUDZKIEGO UCZNIÓW. CZĘŚĆ II: METODOLOGIA I REGULY

STRESZCZENIE

Paradygmaty rozwojowe, bardziej cenione w diagnostyce edukacyjnej niż paradygmaty interwencyjne, wymagają stosowania skal pionowych, rozciągniętych w czasie, i szacowania wartości dodanej jako postępu osiągnięć ucznia lub grupy uczniów w pewnym okresie. Tę wartość można przedstawić w surowych wynikach testu, centylach, staninach, równoważnikach klasy lub jednostkach probabilistycznej skali „teta”.

Równoważniki klasy to średnie liczby lat i miesięcy nauki potrzebne do uzyskania danego poziomu osiągnięć. Łatwo je obliczyć, lecz mogą być mylące. Laicy chcieliby przesunąć uczniów odpowiednio w górę lub w dół drabiny szkolnej, co zrujnowałoby system edukacyjny. Znacznie bardziej statystycznie wyrafinowana jest teoria wyniku zadania (*IRT*), oparta na oszacowaniu zależności wyniku zadania od właściwości stanowiącej zmienną ukrytą. *IRT* może być stosowana do zadań testów umysłowych, pozycji kwestionariuszy, kategorii obserwacyjnych i nawet do informacji biograficznych.

Najsilniej wpływową zmienną kontekstową w diagnostyce edukacyjnej jest pozycja społeczno-ekonomiczna rodziny, warunkująca zarówno aspiracje edukacyjne młodzieży, jak

też wyniki egzaminów szkolnych. Dwa dalsze zjawiska związane z kształceniem szkolnym stanowią: bezradność intelektualna, wywołana systematycznym niepowodzeniem w uczeniu się jednego lub więcej przedmiotów szkolnych, i wyuczona jednostronność, wynikająca z pragnienia zdobycia i utrzymania pozycji najlepszego ucznia.

Większość systemów edukacyjnych dąży do wyrównywania szans edukacyjnych każdego dziecka, zmniejszania różnic między najlepszymi i najslabszymi uczniami oraz do tego, by „żadne dziecko nie zostawało w tyle”. Te słuszne politycznie idee nie ostają się jednak wobec psychologicznego prawa efektu wachlarzowego w postaci wzrostu wariacji osiągnięć wraz ze wzrostem ich średniej w danej populacji.

Słowa kluczowe:

diagnostyka edukacyjna, równoważnik klasy, teoria wyniku zadania, bezradność intelektualna, wyuczona jednostronność, efekt wachlarzowy.