ZARZĄDZANIE PUBLICZNE 3–4(62)2024, s. 83–96
https://doi.org/10.4467/20843968ZP.24.007.22570
www.ejournals.eu/czasopismo/zarzadzanie-publiczne

https://orcid.org/0000-0002-0346-712X
Bartłomiej Walczak
Uniwersytet Warszawski
e-mail: b.walczak@uw.edu.pl

USING LEARNING TRAJECTORIES IN SMALL-SCALE EVALUATION

Abstract

The text is a proposal to apply learning trajectories (also known as progression lines) in educational evaluation. The author embeds learning trajectories in the evolution of the theory of evaluation, showing them as an attempt to combine elements from the post-positivist and phenomenological approaches (in particular, constructivist and transformational evaluation). The learning trajectories are also part of the shift towards micro-evaluation, which follows attempts to move evaluation practices beyond the circle of professional evaluators. The rest of the text presents a review of the literature on the learning trajectories, including areas of application and effectiveness research. In the last section, the author proposes a simple model adapting the learning trajectories to the conceptualization of evaluation.

Keywords: evaluation, self-evaluation, learning trajectories, progressions lines

Abstrakt

Zastosowanie linii rozwoju w mikroewaluacjach

Tekst jest propozycją zastosowania rozwiązania znanego w polskiej literaturze pod nazwą „linie rozwoju" (ang. *progression lines*, *learning trajectories*) w ewaluacji edukacyjnej. Autor umiejscawia linie rozwoju na tle ewolucji teorii ewaluacji jako próbę połączenia elementów pochodzących z nurtu postpozytywistycznego i nurtów fenomenologicznych, w szczególności ewaluacji konstruktywistycznej i transformatywnej. Linie rozwoju wpisują się ponadto w zwrot w stronę mikroewaluacji, związany z przenoszeniem praktyk ewaluacyjnych poza krąg zawodowych ewaluatorów. W dalszej części tekstu zostaje przedstawiony przegląd literatury dotyczącej linii rozwoju z uwzględnieniem obszarów zastosowania i badań nad efektywnością. W ostatniej sekcji autor proponuje prosty model, dostosowujący linie rozwoju do pojęciowości ewaluacji.

Słowa kluczowe: ewaluacja, autoewaluacja, linie rozwoju

Introduction

This paper is intended to propose a practical evaluation solution combining the latest trends in formative assessment and the need for an agile evaluation. The learning trajectories are widely discussed in the literature, applied in assessment, included in the curricula, and tested for impact and validity but are not used in evaluation. In the first section of the text, I would like to embed learning trajectories in the evolution of the evaluation theory and the recent shift toward small-scale evaluation. Then, I will describe the concept of learning trajectories and review the relevant literature. In the last section, the simple model of the use of learning trajectories in evaluation is proposed. I hope that this solution will be useful not only to professional evaluators but also to teachers and educators.

I decided to use the term "learning trajectories" as it was originally used by Martin M. Simon [1995]. It should be noted that the term learning progression is used interchangeably.

A short glimpse into the history of evaluation

The origins of educational evaluation can be traced back to the 18th century when Cambridge chemist William Farish introduced a quantitative grading system. Different grading systems, subsequently developed in many countries, allowed the measurement of student progress and evaluation of the quality of teaching [Hogan, 2007]. However, it was Joseph Rice, who, since 1887, conducted the first educational program evaluation on an impressive sample of 38 thousand students [Guba, Lincoln, 1989; Rice, 1897]. Rice's research was focused on grammar learning. He aimed to improve the learning process by adding advanced programs to the curriculum and enhancing teachers' competencies [Hildreth, 1979; Rice, 1897]. Rice's work marks the beginning of the era of qualitative measurement in evaluation. At the beginning of the 20th century, thousands of psychometric tools for measuring students' competencies were developed [Hildreth, 1979], but the theory's development did not follow measurement. The time for the theory came in the 1940s. Lance Hogan [2007: 5] named this period the "Tylerian Age". Ralph Tyler's work brought evaluation to the next level by providing a new epistemological perspective by using evaluation criteria. It allowed for the further development of specific methodologies. His orientation was positivist, but his understanding of evaluation went beyond test-oriented measurement [Madaus, 2004]. The next milestone has come in 1960s, when Robert Stake [1967] acknowledged the role of valuing in evaluation, contrary to the widespread idea of value-free science. His famous statement naming description and judgement as basic activities in evaluation shaped the unique epistemological perspective of evaluation.

Since the late 1960s and 1970s, evaluation has started to develop into different epistemological orientations and methodologies. New subsets of qualitative methodologies were introduced as alternatives to psychometrics and experimental

designs [Britan, 1978; Wilson, 1977; Wilson and others, 1974]. The main critique of the positivist, quantitative-oriented evaluation was its inability to cover all the aspects of the evaluated projects, as well as the contexts [Koppelman, 1979]. Apart from the methodology, the 1970s qualitative turn revealed some other issues that can be seen as a presage of upcoming new theoretical perspectives, particularly constructivism and transformative approach.

The constructivist approach, which emerged from phenomenologically oriented sociology in the late 1960s, stressed how social actors define reality [cf. Becker, 2009]. There was also a clear influence of the hermeneutic philosophy, with its emphasis on the relativeness of meanings. Two basic constructivist assumptions: people are building their understanding of reality, and those understandings do not necessarily overlap, meet one of the basic observations in evaluation: any program can be seen by different stakeholders from different, sometimes opposed perspectives. Evaluanduum is a space of different subcultures, roles and politics. The program's success largely depends on understanding the unique perspectives of the key stakeholders, therefore it is an evaluator's responsibility to trace the social world complexity through the lenses of social actors [Schwandt, 2000]. Coming back to Stake's emphasis on valuing judgments, the evaluation process is by default embedded in evaluators' set of values; there is still a question of how to catch the complexity of values shared by stakeholders and how to involve them in the evaluation process.

A second question, raised by the constructivists, is put on the banners of a wide and internally differentiated transformative approach [Mertens, 2015]. How can we give a voice to the stakeholders, particularly the unprivileged ones? Can the evaluation be an empowerment tool for the subalterns? Depending on the theoretical origins: neo-Marxist, feminist, postcolonial studies, etc., the exact focus may be different, but emancipatory potential is clear. The transformative approach brought new ethical dilemmas into the stake but also raised the awareness of evaluators of silenced voices and subtle power relationships within the evaluanduum [Fetterman, 1996].

The intention of this short review was not to provide a comprehensive picture of the history of evaluation but to show the key intellectual trends that are the basis for the model presented in a subsequent part of this paper. The trajectory – positivism, phenomenology, constructivism, and transformative approach – does not cover all the theoretical trends in evaluation. In fact, it is not even a trajectory. While the shift from a positivist to a phenomenological approach can be seen as a linear change (with a strong position of post-positivism held up to today), the development of the evaluation theory is multilinear and homogeneous [Mizerek, 2017]. As a Reader probably noticed, I avoid the use of the term "paradigm" [cf. Martens, Dams-O'Connor, Beck, 2006] to describe the distinctive theoretical approaches, as they do not meet the philosophy of science understanding of the term [Kaiser, 2020; Kuhn, 2001]. Since the 1960s, most of the theoretical branches in the evaluation have been developed simultaneously, and theorists have been

arguing but also drawing from others. We definitely cannot talk about a subsequent change of paradigms or completely distinctive approaches. All "paradigms" in evaluation share a lot while maintaining some distinctive differences and focuses. Thus, learning trajectories, combining positivist-like design with a constructivist approach to the learning goals and criteria, and transformative openness for the learners' are not so theoretically doubtful proposal as it may appear at the first glance.

Towards small-scale evaluations

Since the Rice times, educational evaluation was used as a tool for tracing the intervention and quality improvement. Evaluators were mostly academics and professional researchers, with little, if any, space left for teachers. Evaluation projects typically covered units far more bigger than a classroom. There are, however, reasonable arguments towards the wide use of evaluation by teachers and educators on the small unit level. The first argument comes from the policy level. Implementing self-evaluation in different educational systems was one of the instances of decentralization of education systems in many countries. Schools were given more autonomy and responsibility for the quality of education, along with more pressure on accountability. Evaluation and external exams are two basic tools to meet this requirement [Chapman et al., 1996; Varkas, 2022]. The second argument is about the impact of evaluation. There is a bunch of evidence on the impact of external evaluation (namely the school inspection) on schools. Systematic reviews have shown that it changes the way schools operate and influences teachers' practices, but its impact on the learning outcomes is limited [Klerks, 2013; Nelson, Ehren, 2014]. Although research on the impact of self-evaluation is limited, some evidence is available. The literature review by Nelson and coauthors [2015] shows a positive impact on the students' experiences of learning and educational outcomes. Self-evaluation makes teachers more aware of the learning conditions, which influences the overall quality of teaching. Teachers also learn new diagnostic tools [Nelson, Ehren, Godfrey, 2015]. If conducted in teams, self-evaluation increases communication and cooperation between teachers [Kołodziejczyk, Kołodziejczyk, 2015]. Of course, evaluation effectiveness is only one side of the coin. There must be a need for improvement in schools [MacBeath, McGlynn, 2005: 17]. Finally, there is also the third, practical argument. Complex evaluation projects are costly and time-consuming. They require high research and analytical skills, and specific resources. Therefore, it is naïve to expect schools to conduct complex evaluation and growing literature shows how to evaluate on a small-scale [Airasian, Gullickson, 1997; MacBeath, 1999; Robson, 2000]. Small-scale evaluations are typically local, involve single evaluator, and are run within limited time and on limited resources [Robson, 2000]. The model proposed in subsequent sections of this paper fits the small-scale approach.

Learning trajectories

Learning trajectories, also referred to as progression lines [Confrey, Maloney, Nguyen, 2014; Daro, Mosher, Corcoran, 2011], are typically defined as empirically testable hypotheses about learning progress [Corcoran, Mosher, Rogat, 2009]. More specifically, the learning progress is understood as stepwise progress from the basic ideas (lower-anchor) to scientifically based explanations (upper-anchor) [Furtak, 2018]. Trajectories are built upon "big ideas" – central conceptual structures that integrate elements of the curriculum. A good example of such a central concept is natural selection in evolution, which coins together many otherwise unrelated elements, providing a general explanatory framework [Catley, Lehrer, Reiser, 2005]. "Big ideas" are not necessarily very wide domains (like, for example, science or mathematics), as learning trajectories are more practical when developed within more narrow and specific domains [Shepard, 2018].

The term was proposed by Martin M. Simon [1995] as a hypothetical learning trajectory. Giving a rationale embedded from a constructivist perspective, he described hypothetical learning trajectory as the teacher's prediction of the student's learning path. The trajectory assumes regularity and processualness of learning, and its course is anticipated (namely hypothesized) based on the teacher's experience and knowledge. While being a construct, a hypothetical learning trajectory provides a rationale for the teacher's action in the classroom. It consists of three elements:

1. The learning goal that describes a targeted developmental level;
2. Developmental progression – increasing stages leading to the learning goa
3. Instructional activities covering tasks and strategies targeted at developing each level [Baroody et al., 2021; Simon, 1995].

The whole trajectory is subject to change as the components and teacher's knowledge evolve [Simon, 1995]. As Simon himself metaphorically described it:

> Consider that you have decided to sail around the world in order to visit places that you have never seen. One does not do it randomly (e.g. go to France, then Hawaii, then England), but neither is there one set itinerary to follow. Rather, you acquire as much knowledge relevant to planning your journey as possible. You then make a plan. You may initially plan the whole trip or only part of it. You set out sailing according to your plan. However, you must constantly adjust because of the conditions that you encounter. You continue to acquire knowledge about sailing, about the current conditions, and about the areas that you wish to visit. You change your plans with respect to the order of your destinations. You modify the length and nature of your visits as a result of interaction with people along the way. You add destinations that prior to your trip were unknown to you. The path that you travel is your "trajectory". The path that you anticipate at any point in time is your "hypothetical trajectory" [Simon, 1995: 136–137].

In fact, Simon's hypothetical learning trajectory owes a lot to cognitively guided instructions. The research within that approach, conducted in the early

1990s attempted to understand how teachers can support students in moving along the mathematical processes they learned. One of the results was the important role of evidence-based pathways in guiding instruction and supporting teachers [Furtak, 2018]. In the early 2000s, researchers started to document subsequent stages of the observed students' learning process [Catlet et al., 2005]. Soon, the learning trajectories started to be used to track students' progress in different subjects, from biology [Duncan, Rogat, Yarden, 2009; Elmesky, 2013; Wulandari et al., 2019], chemistry [Cooper, Klymkowsky, 2013; Cooper, 2013], geography [Huynh, Solem, Bednarz, 2015; Lane, Bourke, 2019], though languages [Bennett, Deane, Rijn van, 2016; Greaney, Tunmer, 2010], sciences [Chen et al., 2014; Lacy et al., 2014) up to the math [Briggs, Peck, 2015; Clements et al., 2011, 2017, 2020], which is apparently the best developed. Apart from typical school subjects, we can find learning trajectories used in transversal skills measurement [Białek, Swat-Pawlicka, 2022; Rich et al., 2019]. Learning trajectories are applied in national curriculums and validated assessment designs [Callingham, Watson, Oates, 2021; Confrey, Maloney, Corley, 2014; Furtak, 2018]. The linear structure makes learning trajectories a convenient progress measurement tool [Duncan, Hmelo-Silver, 2009].

While well-developed in children's education, particularly in mathematics teaching, the literature on the use of learning trajectories in adult education is limited. Most publications come from research on college and university-level students learning [Reed, Wolfson, 2021; Testa et al., 2019], as well as from studies of teachers' professional development [Bargagliotti, 2020].

The growing literature shows a positive impact of the use of learning trajectories on the learning processes, but the evidence is limited. The most reliable, RCT-based evidence comes from samples of young children. Practically all studies were conducted in the USA and were oriented toward mathematic learning. An experimental study by John Fantuzzo and co-authors [Fantuzzo, Gadsden, McDermott, 2011] on a sample of 1415 young children (between 3–6 years old) revealed growth in learning trajectories group in mathematics and listening skills when controlled for demographics, special needs and language status. Another randomized trial study on preschoolers by Douglas Clements and Julie Sarama [2008] used a three-group design on a random sample of 35 teachers and 276 children, dividing children into intervention, comparison (other curriculum), and control groups. The results show significant positive changes in the classroom environment and teaching practices, as well as large effects on mathematical achievements. A randomized study of 2027 students aged 4-6 from low-income households revealed a higher level of performance in math in the intervention group, and no significant impact on letter and word identification, but differences between groups were fading over time [Clements et al., 2011]. A small-scale (n=16) RCT by Arthur Baroody and co-authors on 3–4-year-olds [Baroody et al., 2021] showed a small significant effect in the intervention group and some implications for the instructional ordering. The results of another two randomized

trials performed by Douglas Clements [Clements et al., 2019, 2021] have proven the role of subsequent steps in learning through the trajectory. The only non-US-based study in this review is a small-scale (n=47) RCT on 4- to 5-year-olds in Turkey [Ceylan, Aslan, 2024] that showed significant improvement in preschoolers' mathematic skills.

The only randomized trial using a sample other than preschoolers I was able to identify was Dana Christensen and Doug Lombardi [2023] study on biological evolution learning and computational thinking. The results from a sample of 2843 high school students show medium to large effects on students' knowledge.

As we can see, learning trajectories have become one of the modern trends in formative assessment, namely "one of the most important assessment design ideas to be introduced in the past decade" [Shepard, Daro, Stancavage, 2013]. The applications are spread in many countries and different subjects, and a high level of validity has been proven [Baroody et al., 2021; Chen et al., 2017; Chen et al., 2016; Gao et al., 2023; Herrmann-Abell, DeBoer, 2018]; however, some serious drawbacks should be noted. The first one is limited evidence on its impact apart from preschoolers learning mathematics. Designing a reliable learning progression requires substantial knowledge of the subject, especially of the research base [Gotwals, 2012], and the possibility of its validation by individual teachers is limited. We can also note that the core idea of "trajectory" assumes linear progression in learning. As it was shown by Amelia Gotwals and Nancy Butler Songer [2010], there is a challenge of a "messy middle". Progress between the trajectory's lowest and highest ends does not have to be linear. Gotwals indicated more challenges: a low transferability of learning trajectories between the subjects and stakeholders, potentially confusing differences between highly specific, "zoomed-in" trajectories and more general, "zoomed-out" ones, and difficulties in delivering a learning trajectory as a ready-to-use product for the teachers [Gotwals, 2012].

The use of learning trajectories in evaluation

As the Reader may have noted, the concept of learning trajectories is an evaluative concept. Learning trajectories allow students to trace their progress, simultaneously giving the teacher feedback about the quality of teaching and learning. By defining the milestones, the evaluation-like criteria are provided, and there is a logic in the longitudinal design similar to designs used in impact evaluation [Bamberger, Rygh, Mabry, 2006]. However, there are practically no applications in evaluation. One example of use I was able to identify was a two-year evaluation of a pilotage program in Poland [Walczak, Wąsowska, 2022]. Learning trajectories (here: progression lines) were designed to measure the impact of intervention focused on improving the pro-innovative skills in primary students, along with multi-sited anthropological evaluation.

Apart from big evaluation projects, learning trajectories can be a useful tool in small evaluations [Robson, 2000], including adults' learning, in particular self-evaluation used by teachers and educators to get feedback on their work. Applying learning trajectories logic into self-evaluation or small-scale external evaluation should increase the quality of learning by providing a more informed structure and giving learners better feedback on their achievements. I would like to propose a four-step procedure:

1. Defining learning goals.
2. Mapping the milestones.
3. Measurement of progress.
4. Reflection.

In the teaching practice, the development of the learning trajectory starts with defining the teaching objectives. Knowing the objectives, necessary skills and competencies can be defined. The goals must be challenging but achievable within the scheduled time and with available resources. Albeit learning trajectory theorists proposed to focus on the "big ideas", I found this concept too imprecise in my evaluation practice. The idea reflected in the learning goal must be "big" enough to allow for mapping subsequent steps in the learning, but the evaluation structure proposed here will work not only for the key ideas within the discipline. Consider a simple example of teaching about bicycle service. At first glance, the ability to change a tube may work as a learning goal. It consists of subsequent activities: removing a tyre and broken tube, inserting a new tube, securing the valve, putting the tyre back, and pumping up the tube. However, the bicyclist must be able to do all the elements to get their bicycle back on the track. Being able to insert a new tube without any pumping skills (ok, I can see the limits of this example) makes the goal unachieved. We cannot see a gradual development of the competencies here. The order of the elements is based on a temporal trajectory, not a competency trajectory. Let's try to widen the learning goal and name it as self-serve of the bicycle. By the self-serve, I understand being capable of performing all the services necessary to ride the whole season. It covers situational skills (e.g. ability to change a broken tube), maintenance skills (e.g. chain lubrication), and regulation skills (e.g. derailleur adjustment or wheel centring). I believe that even with this imperfect example, the Reader can see a trajectory. Some skills are of a basic level and necessary when an emergency (broken tube) occurs, while others (derailleur adjustment) can increase the comfort of riding when there is no bicycle workshop around. Since each element consists of easily definable activities, a time framework and necessary resources can be scheduled. My biking experience says it is a challenging goal but (to some extent) achievable.

The evolution of evaluation, discussed in the first part of this paper, shows the role of stakeholders' involvement. Evaluation has empowerment potential and can provide the space for stakeholders who are not usually heard. This approach requires, however, an openness to hear different voices. Discussing a bicycle self-serve as a learning goal may be oversimplified. But even here, we can

expect differences between the learners that may influence the goal's achievability. Discussing and deciding on the real learning goals will be more demanding. Apart from the differences (educational, gender, ethnic, social class, age etc.) the shared understanding of the goal must be achieved by opening the discussion about goals with the learners.

The learning trajectories should be based on empirically tested evidence, but measurement experts advise a "horizontal coherence" by building the learning trajectory bottom-up with a focus on local jurisdiction or curricula to ensure coherence in curriculum, instruction, assessment and teachers' training [Shepard, 2018]. In many cases, the teacher or educator's experience and knowledge may be enough to project the trajectory. It is easy to design the trajectory for the bicycle self-serve based on the casual biking experience, situating situational skills (changing the tube) on the lower anchor through maintenance (lubricating the chain) up to the upper anchor regulation skills (derailleur adjustment). Adding specific skills requiring more advanced tools (centring the wheel) can open the upper anchor, which helps in projecting future development [Białek, Swat-Pawlicka, 2022]. In education, milestones must be defined according to the knowledge about skill development. Inviting learners to the design process will ensure empowerment (and motivation). From the evaluation point of view, this is a defining stage for criteria. What makes learning trajectories valuable is an emphasis on the linear process of achieving the goals, instead of dichotomic, "achieved" vs "not achieved" distinction.

The data collection process can involve all solutions from the evaluation toolbox, but the minimalistic design of small evaluation affords the simplest solutions. By the definition, evaluation with the learning trajectories uses a longitudinal design. Measurements can be performed with competence tests or learners' self-evaluations of their skills, but tools typical for the teachers' toolbox may also be useful. One of the most obvious solutions is the observation of learners' activities and performance during the learning process. An analysis of the products created during the learning process can bring data on the learners' progress. The minimalism of small evaluation does not stop an evaluator from using triangulated methods and data sources.

Finally, the results of the evaluation, apart from providing the teacher or educator with feedback information, should be discussed with the learners. It would allow the learners to know their positions on the learning trajectories and help close the learning process. Coming back to the bicycle workshop: the self-dependent bicyclist should be aware of their own competencies and challenges they may attempt to overcome on the track.

Conclusion

In the presented papers, I tried to argue for the use of learning trajectories in evaluation by showing their coherence with the current state of the art in evaluation

theory and the shift towards small-scale evaluation. Learning trajectories are recognized as one of the most important designs in the formative assessment and allow for significant increments in the learning process efficiency, being a good symbol of change in assessment since Farish. Adaptation to the evaluation as a kind of meso-structure is straightforward and can bring a more nuanced approach to the evaluation criteria. However, learning trajectories are proven to effectively support teaching and learning. I hope that they may become a part not only of the evaluators but also of the teaching providers' toolbox.

References

Airasian P.W., Gullickson A.R. (1997), *Teacher Self-Evaluation Tool Kit*, Corwin Press, Thousand Oaks.

Bamberger M., Rugh J., Mabry L. (2006), *Real World Evaluation. Working Under Budget, Time, Data, and Political Constraints*, SAGE Publications, Los Angeles.

Bargagliotti A. (2020), Developing learning trajectories for teacher learning: learning trajectories for teachers, *International Journal of Teacher Education and Professional Development*, 3(2), 33–46, https://doi.org/10.4018/IJTEPD.2020070103.

Baroody A., Yilmaz N., Clements D., Sarama J. (2021), Evaluating a basic assumption of learning trajectories: the case of early patterning learning, *Journal of Mathematics Education*, 14(2), 8–32, https://doi.org/10.26711/007577152790071.

Becker H. (2009), *Outsiderzy. Studia z socjologii dewiacji*, transl. O. Siara, Wydawnictwo Naukowe PWN, Warszawa.

Bennett R.E., Deane P., Rijn P.W. van (2016), From cognitive-domain theory to assessment practice, *Educational Psychologist*, 51(1), 82–107, https://doi.org/10.1080/00461520.2016.1141683.

Białek K., Swat-Pawlicka M. (2022), *Diagnoza poziomu kompetencji proinnowacyjnych uczniów*, Centrum Edukacji Obywatelskiej.

Briggs D.C., Peck F.A. (2015), Using learning progressions to design vertical scales that support coherent inferences about student growth, *Measurement: Interdisciplinary Research and Perspectives*, 13(2), 75–99, https://doi.org/10.1080/15366367.2015.1042814.

Britan G. (1978), The place of anthropology in program evaluation, *Anthropological Quarterly*, 51(2), 119–128.

Callingham R., Watson J., Oates G. (2021), Learning progressions and the Australian curriculum mathematics: the case of statistics and probability, *Australian Journal of Education*, 65(3), 329–342, https://doi.org/10.1177/00049441211036521.

Catley K., Lehrer R., Reiser B. (2005), *Tracing a Prospective Learning Progression for Developing Understanding of Evolution* [Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievement], National Academies Committee on Test Design for K-12 Science Achievement.

Ceylan M., Aslan D. (2024), The effect of learning trajectories-based coding education program on preschoolers' mathematical measurement skills, *Education and Information Technologies*, 29(7), 7737–7757, https://doi.org/10.1007/s10639-023-12107-7.

Chapman J., Boyd W., Lander R., Reynolds D. (1996), *The Reconstruction of Education. Quality, Equality and Control*, Cassell, London.

Chen F., Zhang S., Guo Y., Xin T. (2017), Applying the rule space model to develop a learning progression for thermochemistry, *Research in Science Education*, 47(6), 1357–1378, https://doi.org/10.1007/s11165-016-9553-7.

Chen J., Gotwals A.W., Anderson C., Reckase M. (2016), The influence of item formats when locating a student on a learning progression in science, *International Journal of Assessment Tools in Education*, 3(2), 101–122, https://doi.org/10.21449/ijate.245196.

Chen R.F., Eisenkraft A., Fortus D., Krajcik J., Neumann K., Nordine J., Scheff A. (eds.) (2014), *Teaching and Learning of Energy in K – 12 Education*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-05017-1.

Christensen D., Lombardi D. (2023), Biological evolution learning and computational thinking: Enhancing understanding through integration of disciplinary core knowledge and scientific practice, *International Journal of Science Education*, 45(4), 293–313, https://doi.org/10.1080/09500693.2022.2160221.

Clements D.H., Fuson K.C., Sarama J. (2017), The research-based balance in early childhood mathematics: A response to Common Core criticisms, *Early Childhood Research Quarterly*, 40, 150–162, https://doi.org/10.1016/j.ecresq.2017.03.005.

Clements D.H., Sarama J. (2008), Experimental evaluation of the effects of a research-based preschool mathematics curriculum, *American Educational Research Journal*, 45(2), 443–494, https://doi.org/10.3102/0002831207312908.

Clements D.H., Sarama J., Baroody A.J., Joswick C. (2020), Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction, *ZDM*, 52(4), 637–648, https://doi.org/10.1007/s11858-019-01122-z.

Clements D.H., Sarama J., Baroody A.J., Joswick C., Wolfe C.B. (2019), *Evaluating the efficacy of a learning trajectory for early shape composition, American Educational Research Journal*, 56(6), 2509–2530, https://doi.org/10.3102/0002831219842788.

Clements D.H., Sarama J., Baroody A.J., Kutaka T.S., Chernyavskiy P., Joswick C., Cong M., Joseph E. (2021), Comparing the efficacy of early arithmetic instruction based on a learning trajectory and teaching-to-a-target, *Journal of Educational Psychology*, 113(7), 1323–1337, https://doi.org/10.1037/edu0000633.

Clements D.H., Sarama J., Spitler M.E., Lange A.A., Wolfe C.B. (2011), Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial, *Journal for Research in Mathematics Education*, 42(2), 127–166, https://doi.org/10.5951/jresematheduc.42.2.0127.

Confrey J., Maloney A., Nguyen K. (2014), Learning trajectories in mathematics [in:] A. Maloney, J. Confrey, K. Nguyen (eds.), *Learning Over Time: Learning Trajectories in Mathematics Education* (XI–XXII), Information Age Publishing, Charlotte.

Confrey J., Maloney A.P., Corley, A.K. (2014), Learning trajectories: a framework for connecting standards with curriculum, *ZDM*, 46(5), 719–733. https://doi.org/10.1007/s11858-014-0598-7.

Cooper M., Klymkowsky M. (2013), Chemistry, life, the universe, and everything: a new approach to general chemistry, and a model for curriculum reform, *Journal of Chemical Education*, 90(9), 1116–1122, https://doi.org/10.1021/ed300456y.

Cooper M.M. (2013), Chemistry and the next generation science standards, *Journal of Chemical Education*, 90(6), 679–680, https://doi.org/10.1021/ed400284c.

Corcoran T.B., Mosher F.A., Rogat A. (2009), *Learning Progressions in Science: An Evidence-Based Approach to Reform*, CPRE Research Reports.

Daro P., Mosher F., Corcoran T. (2011), *Learning Trajectories in Mathematics. A Foundation for Standards, Curriculum, Assessment, and Instruction*, Consortium for Policy Research in Education.

Duncan D.G., Hmelo-Silver C.E. (2009), Learning progressions: aligning curriculum, instruction, and assessment, *Journal of Research in Science Teaching*, 46(6), 606–609.

Duncan R.G., Rogat A.D., Yarden A. (2009), A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades, *Journal of Research in Science Teaching*, 46(6), 655–674, https://doi.org/10.1002/tea.20312.

Elmesky R. (2013), Building capacity in understanding foundational biology concepts: a K-12 learning progression in genetics informed by research on children's thinking and learning, *Research in Science Education*, 43(3), 1155–1175, https://doi.org/10.1007/s11165-012-9286-1.

Fantuzzo J.W., Gadsden V.L., McDermott P.A. (2011), An integrated curriculum to improve mathematics, language, and literacy for Head Start children, *American Educational Research Journal*, 48(3), 763–793, https://doi.org/10.3102/0002831210385446.

Fetterman D. (1996), Empowermental evaluation: an introduction to theory and practice [in:] D. Fetterman, S. Kaftarian, A. Wandersman (eds.), *Empowermental Evaluation. Knowledge and Tools for Self-Assessment and Accountability*, SAGE, Thousand Oaks–London–New Dheli.

Furtak E. (2018), *Supporting Teachers' Formative Assessment Practice with Learning Progressions*, Routledge, London.

Gao Y., Zhai X., Bae A., Ma W. (2023), Rasch-CDM: applying rasch and cognitive diagnosis models to assess learning progression [in:] X. Liu, W. Boone (eds.), *Advances in Applications of Rasch Measurement in Science Education* (97–122), Springer Nature, Cham.

Gotwals A.W. (2012), Learning progressions for multiple purposes. Challenges in using learning progression [in:] A.W. Gotwals, A. Alonzo (eds.), *Learning Progressions in Science. Current Challenges and Future Directions* (461–472), Sense Publishers, Rotterdam.

Gotwals A.W., Songer N.B. (2010), Reasoning up and down a food chain: using an assessment framework to investigate students' middle knowledge, *Science Education*, 94(2), 259–281, https://doi.org/10.1002/sce.20368.

Greaney K., Tunmer W.E. (2010), The literacy learning progressions and the reading and writing standards: some critical issues, *Kairaranga*, 11(2), 23–27.

Guba E., Lincoln Y. (1989), *Fourth Generation Evaluation*, SAGE, Newbury Park–London–New Dheli.

Herrmann-Abell C.F., DeBoer G.E. (2018), Investigating a learning progression for energy ideas from upper elementary through high school, *Journal of Research in Science Teaching*, 55(1), 68–93, https://doi.org/10.1002/tea.21411.

Hildreth G. (1979), *Bibliography of Mental Tests and Rating Scales*, The Psychological Corporation, New York.

Hogan L. (2007), The historical development of program evaluation: exploring past and present, *Online Journal for Workforce Education and Development*, 2(4), 1–14.

Huynh N.T., Solem M., Bednarz S.W. (2015), A road map for learning progressions research in geography, *Journal of Geography*, 114(2), 69–79, https://doi.org/10.1080/00221341.2014.935799.

Kaiser D. (2020), The paradigm in the philosophy of science [in:] *Economic Theory in the 21st Century: Towards a Renewed Understanding of Money and Capital from a System-wide Perspective*, Springer Fachmedien Wiesbaden, 3–4, https://doi.org/10.1007/978-3-658-30639-7_1.

Klerks M. (2013), *The Effect of School Inspections: A Systematic Review*, UCL Institute of Education, London.

Kołodziejczyk J., Kołodziejczyk J. (2015), Ewaluacja własnej pracy nauczycieli – obszary, rozumienie, współpraca, *Forum Oświatowe*, 27(2), 115–130.

Koppelman K. (1979), The explication model: An anthropological approach to program evaluation, *Educational Evaluation and Policy Analysis*, 1(4), 59–64.

Kuhn T.S. (2001), *Struktura rewolucji naukowych*, przeł. H. Ostromęcka, Fundacja Aletheia, Warszawa.

Lacy S., Tobin R.G., Wiser M., Crissman S. (2014), Looking through the energy lens: a proposed learning progression for energy in grades 3–5 [in:] R.F. Chen, A. Eisenkraft, D. Fortus, J. Krajcik, K. Neumann, J. Nordine, A. Scheff (eds.), *Teaching and Learning of Energy in K – 12 Education* (241–265), Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-05017-1_14.

Lane R., Bourke T. (2019), Assessment in geography education: a systematic review, *International Research in Geographical and Environmental Education*, 28(1), 22–36, https://doi.org/10.1080/10382046.2017.1385348.

MacBeath J. (1999), *Schools Must Speak for themselves. The Case for School Self-Evaluation*, Routledge, London.

MacBeath J., McGlynn A. (2005), *Self-Evaluation. What's in It for Schools?*, Routledge, London.

Madaus G. (2004), Ralph W. Tyler's contribution to program evaluation [in:] M. Alkin (ed.), *Evaluation Roots. Tracing Theorists' Views and Influences* (69–79), SAGE Publications, Thousand Oaks–London–New Dheli.

Martens M.P., Dams-O'Connor K., Beck N.C. (2006), A systematic review of college student-athlete drinking: Prevalence rates, sport-related factors, and interventions, *Journal of Substance Abuse Treatment*, 31(3), 305–316, https://doi.org/10.1016/j.jsat.2006.05.004.

Mertens D. (2015), *Research and Evaluation in Education and Psychology. Integrating Diversity With Quantitative, Qualitative, and Mixed Methods*, SAGE, Thousand Oaks–London–New Dheli.

Mizerek H. (2017), *Ewaluacja edukacyjna. Interdyskursywne dialogi i konfrontacje*, Oficyna Wydawnicza „Impuls", Kraków.

Nelson R., Ehren M. (2014), *Review and synthesis of evidence on the (mechanisms of) impact of school inspection*, UCL Institute of Education, London, http://schoolinspections.eu/impact/wp-content/plugins/download-monitor/download.php?id=29.

Nelson R., Ehren M., Godfrey D. (2015), *Literature Review on Internal Education*, University College London Institute of Education, London.

Reed C.R., Wolfson A.J. (2021), Are learning progressions a useful pedagogical tool for instructors?, *Journal of College Science Teaching*, 51(2), 32–40. https://doi.org/10.1080/0047231X.2021.12290546.

Rice J.M. (1897), The futility of the spelling grind, *The Forum*, 409–419.

Rich K.M., Strickland C., Binkowski T.A., Franklin D. (2019), A K-8 debugging learning trajectory derived from research literature, *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 745–751, https://doi.org/10.1145/3287324.3287396.

Robson C. (2000), *Small-scale Evaluation. Principles and Practice*, SAGE Publications, Thousand Oaks.

Schwandt T.A. (2000), Three epistemological stances for qualitative inquiry: Interpretativis, hermeneutics, and social construtionism [in:] N. Denzin, Y. Lincoln (eds.), *Handbook of Qualitative Research* (189–214), SAGE, Thousand Oaks–London–New Dheli.

Shepard L. (2018), Learning progressions as tools for assessment and learning*, Applied Measurement in Education*, 31(2), 165–174.

Shepard L., Daro, P., Stancavage F. (2013*), The Relevance of Learning Progressions for NAEP (NAEP Validity Studies)*, National Center for Education Statistics, Washington.

Simon M. (1995), Reconstructing mathematics pedagogy from a constructivist perspective, *Journal for Research in Mathematics Education*, 26(2), 114–145.

Stake R.E. (1967), The countenance of educational evaluation, *Teachers College Record: The Voice of Scholarship in Education*, 68(7), 1–15, https://doi.org/10.1177/016146816706800707.

Testa I., Capasso G., Colantonio A., Galano S., Marzoli I., Scotti Di Uccio U., Trani F., Zappia A. (2019), Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and Rasch analysis, *International Journal of Science Education*, 41(3), 388–417, https://doi.org/10.1080/09500693.2018.1556414.

Varkas M. (2022), School self-evaluation: a bureaucratic obligation or an opportunity for school development and education quality improvement, *IJAEDU – International E-Journal of Advances in Education*, 8(22), 92–101, https://doi.org/10.18768/ijaedu.1070055.

Walczak B., Wąsowska M. (2022), *Raport z ewaluacji końcowej projektu pilotażowego „Szkoła dla innowatora"*, Centrum Edukacji Obywatelskiej, Warszawa.

Wilson S. (1977), The use of ethnographic methods in educational evaluation, *Human Organization*, 36(2), 200–203, https://doi.org/10.17730/humo.36.2.81682045808j3287.

Wilson S., Sheppers E., Wilson T., Moore D., Wilson P. (1974), *The Use of Ethnography in Educational Evaluation*, Center for New Schools, Chicago.

Wulandari A., Maridi, Sutarno, Ramli M. (2019), Learning progression on conceptual understanding of biology: a systematic review, *AIP Conference Proceedings*, 2194(1), 020142, https://doi.org/10.1063/1.5139874.