
RECENZJE

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for educational and psychological testing*, American Educational Research Association, Washington 2014, 240 s.

W drugiej połowie ubiegłego roku Amerykańskie Towarzystwo Badań Edukacyjnych (*American Educational Research Association – AERA*) opublikowało kolejną edycję *Standardów edukacyjnego i psychologicznego testowania* (AERA i in. 2014). Wersję tę przygotował Wspólny Komitet ds. Standardów Pomiaru Edukacyjnego i Psychologicznego (*Joint Committee on the Standards for Educational and Psychological Testing*) trzech towarzystw naukowych: Amerykańskiego Towarzystwa Psychologicznego (*American Psychological Association – APA*), Narodowej Rady do spraw Pomiarów Edukacyjnych (*National Council on Measurement in Education – NCME*) oraz AERA. W składzie Komitetu, powołanego w 2008 roku, znaleźli się: Barbara S. Plake (jako współprzewodnicząca), Laurens L. Wise (współprzewodniczący), Linda L. Cook, Fritz Drasgow, Brian T. Gong, Laura S. Hamilton, Jo-Ida Hansen, Joan L. Herman, Michael T. Kane, Michael J. Kolen, Antonio E. Puente, Paul R. Sackett, Nancy T. Tippins, Walter D. Way oraz Frank C. Worrell.

Nowa edycja *Standardów* zawiera 249 standardów profesjonalizmu w zakresie pomiaru edukacyjnego i psychologicznego, uporządkowanych w trzech częściach. Część pierwsza – *Podstawy* – obejmuje 68 standardów związanych z problematyką trafności (1), rzetelności (2) oraz stroniczości (3) testów. W części *Operacje* zamieszczono 119 standardów dotyczących projektowania i konstrukcji testu (4), wyników, skal i norm (5), przeprowadzania testu, punkto-

wania zadań, raportów i interpretacji (6), dokumentacji testu (7), a także praw i obowiązków jego użytkowników (8–9). W *Zastosowaniach testów* znalazły się standardy rekomendowane w dziedzinach testów psychologicznych (10), zawodowych (11), edukacyjnych (12) oraz ewaluacji (13). W każdym rozdziale dwóch pierwszych części wskazano rekomendację nadrzędną, rozwiniętą w standardach jej podporządkowanych, z których każdy opatrzone wielozdaniowym komentarzem, i poprzedzoną kilkustronicowym wprowadzeniem teoretycznym. W części trzeciej zastosowano ten sam schemat kompozycyjny – z wyjątkiem nadrzędnych standardów, których nie wyodrębniono. Teoretyczny wymiar *Standardów* dopełniono obszernym słownikiem oraz indeksem zastosowanych terminów pomiarowych, ułożonymi w końcowych partiach publikacji.

Wśród wspomnianych dziewięciu rekomendacji nadrzędnych znalazły się między innymi standardy formułowania każdej zamierzonej interpretacji wyników testu, przedstawiania – w odniesieniu do proponowanego zastosowania testu – odpowiednich dowodów jego trafności, rzetelności, bezstronności (*fairness*), minimalizowania – w trakcie procesu konstruowania i użytkowania testu – wariacji irrelewantnej względem konstruktów, udostępniania dokumentacji powstałej podczas prac konstruktorskich, precyzyjnego określania procedur administrowania testu, punktowania zadań, tworzenia raportów i interpretowania wyników, poszanowania określonych praw przystępujących do testu oraz przejmowania odpowiedzialności za właściwe użytkowanie testu przez jego użytkowników. Tradycyjnie rekomendacjom powyższym, podobnie jak towarzyszącym im standardom, nadano formę zdań powinnościowych, na przykład nadrzędna formuła podrozdziału *Standardy rzetelności/precyzji* otrzymała postać: „Odpowiednie dowody rzetelności/precyzji powinny być udostępnione na potrzeby interpretacji związanej z każdym zamierzonym użyciem wyniku” (standard 2.0), zaś jeden z 20 uszczegółwiających ją standardów: „Standardowy błąd pomiaru [...] powinien być podany w jednostkach każdego komunikowanego wyniku” (2.13).

Zasygnalizowana powyżej priorytetyzacja standardów oraz inwersja kompozycyjna, polegająca między innymi na umiejscowieniu w *Podstawach* standardów powiązanych z problematyką bezstronności testu, w poprzedniej – również strukturalnie trójdzielnej – edycji zamieszczonych w jej czterorozdziałowej części drugiej, zatytułowanej: *Bezstronność testowania*, stanowią dwie z trzech istotnych modyfikacji wprowadzonych w piątym już wydaniu *Standardów*. Trzecią współtworzą substancjalne oraz stylistyczne zmiany w tekście przejętym z edycji wcześniejszej. Te pierwsze dotyczą przede wszystkim wspomnianej już kwestii bezstronności praktyk pomiarowych. Obejmują także problemy związane z roz-

wojem technologii informacyjnej oraz złożoności towarzyszące użytkowaniu testów i ich wyników przez podmioty zainteresowane zewnętrzną ewaluacją edukacji. Do ujednoczenia terminologii oraz stylu prowadzą pozostałe zmiany w tekście *Standardów*.

Adresatami w ten sposób zaktualizowanego kodeksu profesjonalizmu w dziedzinie pomiaru edukacyjnego i psychologicznego pozostają nie tylko twórcy, lecz także użytkownicy testów. *Standardy* stanowią bowiem zbiór kryteriów oceny wartości testu, przydatny podczas refleksji i dyskusji nad stanem praktyki pomiarowej. W zamyśle pełnić mogą również funkcję dydaktyczną, nie wyłącznie w amerykańskim kręgu kulturowym. O roli, jaką rzeczywiście odgrywają, wspominają między innymi Plake i Wise (2015), podkreślając, iż jak dotąd *Standardom* nie nadano mocy prawnej, i wskazując potencjalne nadużycia związane z ich użytkowaniem, na przykład przedkładanie litery *Standardów* nad sąd oparty na wiedzy.

Bibliografia

- Plake B.S., Wise L.L. 2014. *What is the role and importance of the revised AERA, APA, NCME Standards for educational and psychological testing?*, „Educational Measurement: Issues and Practice” (Wydanie specjalne: *The AERA/APA/NCME Standards for Educational and Psychological Testing*), t. 33, nr 4, s. 4–12.

Piotr Maciej Skorupiński