

KODEX DOBRÉ PRAXE V JAZYKOVÉM TESTOVÁNÍ - A PROČ?

CODE OF BEST PRACTICE IN LANGUAGE TESTING: DO WE NEED IT?

Martina Hulešová

Abstrakt

Spravedlivý přístup k testovaným je nezbytnou podmínkou etického použití jakéhokoli testovacího nástroje. Tvůrce a poskytovatel testu sdílejí zodpovědnost za kvalitu testu a měli by předkládat důkazy o validitě interpretací výsledků a rozhodnutí. Příspěvek se zaměří na otázky, které se v této souvislosti v jazykovém testování objevují, a to z historické perspektivy i z pohledu současného stavu. Na příkladech etických kodexů a principů dobré praxe naznačíme, jaké cesty mohou vést ke spravedlivému testování, a položíme si otázky týkající se potřeby etického kodexu a principů dobré praxe v ČR.

Klíčová slova

test, validita, dobrá praxe, kodex, standard, princip

Abstract

Fairness is a basic condition for ethical use of any assessment tool. Both test developers and test users are responsible for test quality and they should provide evidence about the validity of the interpretation and consequences of test results. This article focuses on issues related to test fairness and validity, from the historical perspective and in the light of current approaches. Examples of existing Codes, Principles and Standards will be discussed and questions about the usefulness of a local, Czech ethical code will be discussed.

Key words

test, validity, good practice, code, standard, principle

Úvod

V textu se zamýšlíme nad účelem a využitím kodexů, standardů a principů dobré praxe (nejen) v jazykovém testování, nad otázkou, zda, případně proč jsou potřebné, a také nad tím, zda by i Česká republika měla mít svůj vlastní lokální standard nebo kodex upravující praxi v jazykovém testování. Tento text je výzvou k přemýšlení a diskusi, nemá žádný zobecňující závěr, nicméně téma a obsah příspěvku mohou naznačovat, že by kodex dobré praxe nebo podobný dokument v České republice nemusel být zcela nepotřebný.

Východiska

Výchozím bodem pro úvahu o jakémkoli testování by měl být koncept validity. Validita v postmoderním pojetí není izolovatelnou vlastností, ani vlastností testu či testování. Je to koncept, který je abstraktní a přitom přítomný ve všech krocích, které vedou k tvorbě, použití testu a k interpretaci výsledku a uplatnění rozhodnutí činěných na základě této interpretace. Validitu je třeba budovat, nikoli hledat.

Pokud účel testování definujeme jako vyvozování závěrů o výkonu vzhledem k předem danému kritériu, musíme být schopni prokázat (a podpořit důkazy) smysluplnost závěrů a rozhodnutí o testovaných, která činíme na základě výkonů v testu (McNamara, 2000, s. 11).

Je zřejmé, že validace má lokální, kontextově podmíněný charakter, protože kvalitní testovací nástroj je vždy vytvářen ke konkrétnímu účelu pro konkrétní skupinu testovaných a s konkrétní představou o použití jazyka v cílové oblasti (Bachman a Palmer, 1996). Zároveň to však implicitně znamená, že bude existovat široká škála do různé míry lokalizovaných testovacích nástrojů, což může mít za důsledek také problém v rozhodování o vhodnosti a kvalitě nabízených nástrojů.

Vytkneme-li si jako základní principy testování spravedlivost a užitečnost a ptáme-li se, jak jich dosáhnout, pak již začínáme uvažovat o dobré praxi: co to je, jak jí dosahovat a dokazovat. Tyto úvahy o dobré praxi nás dovedou k úvahám o tom, co ještě lze za spravedlivý a užitečný, kvalitní nástroj považovat, a v jakých situacích jím být přestává. A zde již hovoříme o standardech, resp. o minimálních standardech.

Validní nebo spolehlivý nástroj? A etika?

Existuje mnoho teoretických rámců pro validaci testovacích nástrojů. V jazykovém testování patří mezi často citované model Kaneův (Kane, 1992), model Bachmana a Palmera (1996) nebo Weirův sociokognitivní model (2005). Důležité ale je, že se více či méně shodnou v základních aspektech, jimž by měla být věnována pozornost, a také v tom, že validace nástroje začíná již při úvahách o samotné potřebnosti tohoto nástroje, tedy ještě předtím, než vznikne.

Z historické perspektivy vývoje institucionálního jazykového testování od počátku 20. století lze shrnout, že v průběhu doby a na různých místech byla zprvu akcentována buď reliabilita, nebo validita. Psychometrický směr převažující v USA, veden zejména snahou o efektivní selekci (rekrutů, úředníků, zaměstnanců, studentů apod.) a efektivní přidělování finančních prostředků, spíše inklinoval k větší váze reliability, tedy spolehlivosti a replikovatelnosti měření. Bylo akcentováno to, JAK spolehlivé a objektivní je měření. Spolsky (1995) tento systém nazýval *a factory system*. Naopak testování evropské, rozvíjené zejména ve Velké Británii, dlouho psychometriku a objektivizaci odmítalo; v úsilí o komplexnost testování, důraz na to, CO se testuje, tedy na obsah (konstruktovou validitu) a spjatost výuky, kurikula, a upozadovalo některé aspekty související s reliabilitou testovacích nástrojů. Spolsky toto opět trefně označuje, tentokrát jako *a cottage industry* (Spolsky, 1995). Oba proudy se postupně sblížovaly, především od 70. let 20. století, a to i v souvislosti s profesionalizací oboru jazykového testování, vznikem platform pro diskusi (pravidelné konference, specializované časopisy apod.) a také díky nástupu komunikačního přístupu v jazykové výuce. V téže době lze také vysledovat první systematictější zmínky o etice a spravedlivosti v testování.

Aniž bychom hodnotili, lze říci, že validita, reliabilita i etika nebo spravedlivost v testování jsou v současné době důležitými koncepty, s nimiž různým způsobem operuje (nebo by měla operovat) každá zkouška, zejména tzv. zkoušky vysoké důležitosti (high-stakes tests), a to i v České republice.

Kodexy, principy, standardy

V závěru předchozího oddílu jsme zmínili, že validitou, reliabilitou a etikou se zabývají mnozí poskytovatelé a tvůrci zkoušek. Jak to však dělají? Kladou na jednotlivé aspekty stejnou váhu? Nebo něco akcentují? Proč tomu tak je? Interpretují své zkoušky správně? Co je to „správná interpretace“? Interpretují zkoušky srovnatelně s jinými? Je míra, váha a pojetí např. spravedlivosti nebo validity stejné, nebo přinejmenším porovnatelné? Z těchto několika málo otázek vidíme, že stav věcí není tak prostý, jak by se mohlo zdát.

Možná právě proto se i v jazykovém testování začaly ve druhé polovině 20. století objevovat snahy dát různým pohledům společnou perspektivu a mimo jiné počínají etické kodexy, principy dobré praxe a standardy.

Etickým kodexem obvykle rozumíme systematicky zpracovaný soubor obecných i konkrétních norem/předpisů, který nějakým způsobem vymezuje vztahy mezi členy určité komunity a jejich jednání. Etický kodex bývá často považován za nástroj profesní sebedefinice. Členové komunity se k tomuto kodexu hlásí, ne vždy je ale jeho dodržování závazné a ne vždy existují a/nebo jsou uplatňovány sankce za nedodržení. Základním posláním etických kodexů je morální kultivace komunity a stanovení jakýchsi hranic či mantinelů pro ty, kteří se k jejich dodržování přihlásili. Dávají také signál druhé straně, tedy těm, kteří budou s členy komunity přicházet do styku v nejrůznějších situacích a rolích – jako pacienti, klienti, čtenáři apod. Jako

příklad lze uvést kodexy lékařské, právnické, novinářské apod., a v neposlední řadě také kodexy kultivující jednání a postupy komunity jazykových testerů.

Na rozdíl od etických kodexů, které bývají spíše obecněji formulované a stručné, principy popisující dobrou nebo doporučeníhodnou praxi bývají konkrétnější nebo praktičtější zaměřené. Jsou jakýmsi myšlenkovým východiskem, zásadou nebo souborem hodnot a nastavují jakousi povinnost nebo vědomí povinnosti v rámci komunity, která tyto principy formuluje. Zároveň mohou fungovat i jako zásadní nástroj poskytující (jazykovým testerům) oporu při dosahování i dokazování kvalit (testovacích nástrojů) a také mohou napomoci k větší profesionalizaci testerů samotných (Thrasher, 2001). Standardy, nebo dokonce tzv. minimální standardy (viz níže), pak fungují jako velmi konkrétní norma nebo referenční materiál.

Vlastnosti funkčních dokumentů

Jak bylo výše řečeno, kodexy, principy nebo standardy ovlivňují (nebo mohou ovlivňovat) chování. Zároveň je však třeba, aby umožňovaly do jisté míry i flexibilitu a byly lokálně adaptovatelné na ne vždy ideální kontext, neboť naplňování kodexů (principů a standardů) není jen individuální záležitostí, ale souvisí i s podmínkami a kontextem (Jia, 2009). Principy a standardy jsou funkční tehdy, jsou-li splnitelné, akceptovatelné, internalizované, jednoduché a univerzální. Jejich implementaci a akceptaci napomáhá, mohou-li se opřít o analýzu potřeb nebo případové studie. Jsou naplňovány, pokud jsou neustále v oběhu, a naopak, jsou-li v oběhu, znamená to, že jsou využívány nebo přinejmenším reflektovány.

Reflexe principů a standardů v (nejen jazykovém) testování

Pravděpodobně nejčastěji citovaným a nejuznávanějším dokumentem jsou Standardy pro pedagogické a psychologické testování (dále Standardy), (AERA, APA, NCME, 1999), které se vyvíjely již od roku 1954. Definiují zásady dobré praxe v psychologickém a pedagogickém testování a k nim se odkazují nebo z nich čerpají další publikace. V průběhu doby takto např. vznikly *The ETS Standards for Quality and Fairness* (1987), které jsou vlastně komentovanými Standardy aplikovanými v kontextu organizace ETS (*Educational Testing Service*), poskytující mimo jiné zkoušky TOEFL. V Kanadě na základě Standardů vznikl již v roce 1988 *The Code of Fair Testing Practices in Education*, jenž byl dílem *Joint Committee on Testing Practices* a *Canadian Psychological Association*.

Podívejme se nyní konkrétně na tři příklady kodexů a dokumentů ze současné praxe v jazykovém testování. Prvním z nich je *ILTA* (International Language Testing Association) *Code of Ethics*, jež představuje devět principů s anotacemi; na tento dokument navazují *Guidelines for Practice*, jež velmi konkrétně vymezují povinnosti tvůrce testu, podobu administrace testu a také vymezuje práva a povinnosti testovaných. Kodex stanovuje míru uspokojivého etického chování či jednání jazykových testerů a vychází přitom z principů pozitivního účinku, spravedlnosti a respektu k autonomii a občanské společnosti. V anotacích se dokonce ozřejmuje, jaké sankce mohou nastat za nedodržení principů, včetně odejmutí členství v ILTA, a to na základě podnětu etické komise ILTA.

V evropském kontextu dále existují dvě profesní organizace sdružující jazykové testery, a to ALTE (Association of Language Testers in Europe) a EALTA (European Association for Language Testing and Assessment). EALTA operuje s dokumentem *Guidelines for Good Practice in Language Testing & Assessment*, který je od ostatních odlišný zejména tím, že má formu otázek. Tyto otázky se týkají testování, hodnocení a vývoje testů a obrací se explicitně ke třem oblastem a skupinám: k těm, kteří se zabývají přípravou učitelů v testování a hodnocení, dále míří do oblasti školního testování a hodnocení a dále k těm, kteří pracují v oblasti vývoje testů na národní nebo institucionální úrovni. Pro všechny tyto skupiny pak doplňují všeobecné principy jako respekt ke studentům/testovaným, zodpovědnost, spravedlivost,

reliabilita, validita a spolupráce s těmi, jichž se testování týká. Dodržování těchto pokynů je ponecháno na morální odpovědnosti každého ze členů EALTA a nejsou stanoveny explicitně sankce za jejich nedodržení.

Posledním příkladem jsou dokumenty ALTE, které jsou součástí systému vnitřního auditu, které ALTE provádí u zkoušek poskytovaných svými členy nebo zájemci o členství, jehož úspěšné absolvování je podmínkou pro plné členství v této organizaci. Takovým auditem procházejí pravidelně např. také Certifikované zkoušky z češtiny pro cizince poskytované Ústavem jazykové a odborné přípravy Univerzity Karlovy. *Code of Practice a Principles of Good Practice for ALTE Examinations* pokrývají tyto aspekty: vývoj zkoušek, interpretace výsledků, spravedlivost testování a komunikace s uživateli výsledků zkoušek, a to z perspektivy tvůrců zkoušek i z perspektivy těch, kteří budou zkoušky používat. Kodex a principy jsou pro členy závazné a jejich naplňování je v rámci auditu vztaženo k 17 minimálním standardům (Minimal Standards). Dokumenty vycházejí explicitně z výše zmíněného Bachmanova a Palmerova modelu užitečnosti testu (Bachman a Palmer, 1996) a ze Standardů (AERA, APA, NCME, 1999), přičemž minimální standardy jsou popisem toho, co musí být naplněno a prokázáno, s respektem k lokálnímu kontextu zkoušek. Obsah minimálních standardů pokrývá pět základních fází nakládání s testem:

- a) *Konstrukce testu* (teoretický konstrukt, účel a vhodnost pro kontext a populaci, kritéria pro výběr a trénink tvůrců testu, posuzovatelů apod., důkaz o paralelnosti nebo ekvivalenci testových verzí, případně i důkaz o přiřazení k externímu referenčnímu systému).
- b) *Administrace a logistika* (mj. též otázky bezpečnosti testů a ochrany osobních údajů a podpora kandidátů se speciálními potřebami)
- c) *Zpracování dat a vyhodnocení* (přesnost a reliabilita vyhodnocení, včetně dokumentace, způsob výpočtu)
- d) *Analýza testu* (vhodnost a dostatečnost použitých metod statistického zpracování, včetně položkové analýzy a analýzy charakteristik testu)
- e) *Komunikace s uživateli výsledků zkoušek* (způsob, obsah, dostatečnost, rychlost, srozumitelnost)

Závěrečná otázka: Potřebujeme v ČR vlastní kodex dobré praxe?

Doposud jsme představili dokumenty, k jejichž dodržování se dobrovolně přihlašují ti, kteří mají zájem stát se členy některé z uvedených organizací. Ze zkušenosti ÚJOP UK vyplynulo, jak mnoho přispívá zamýšlení se nad standardy a principy k tomu, aby se neustále zvyšovala kvalita služeb, které ÚJOP UK poskytuje, a jaké pozitivní důsledky to má jak směrem dovnitř organizace, tak i navenek, směrem k uživatelům zkoušek, z nichž na prvním místě je třeba jmenovat samotné kandidáty.

Místo závěru bychom proto položili otázky, které by čtenáře směřovaly k reflexi toho, jaký je současný stav nejen jazykového testování v ČR a kam bychom chtěli směřovat:

Pokud by naše odpověď na otázku, zda v ČR potřebujeme vlastní kodex dobré praxe, nebyla jasné NE, pravděpodobně si budeme klást otázky další:

- Komu/k čemu by měl sloužit?
- Kdo by měl/mohl být autorem?

Pokud jsme až dosud myšlenku lokálního kodexu nezavrhlí, můžeme se ptát dále:

- Jaký by byl mechanismus fungování “českého” kodexu?
- Měl by jej někdo „vlastnit“, „spravovat“?
- Měl by vzniknout z iniciativy shora nebo dohodou různých subjektů?

- Pokud by existoval, jak se lze k němu „hlásit“?
- Proč by mělo být výhodné se k dodržování přihlásit?
- Měl by být závazný?
- Kdo by měl dohlížet na dodržování?
- Jaké by byly sankce za nedodržení?
- Měly by existovat institucionální nebo jiné varianty nebo jedna univerzální verze?
- Měla by existovat jen profesní verze, nebo i verze pro uživatele výsledků zkoušek?
- Jak by měl být formulován?
- Měl by obsahovat příklady dobré praxe?
- A příklady špatné/nevhodné praxe?
- Měl by být stanoven minimální standard?

Ve výčtu bychom mohli pokračovat, avšak zásadní otázka zazněla již v nadpisu závěrečného oddílu: **potřebujeme v ČR vlastní kodex dobré praxe?**

Literatura:

BACHMAN, L.F. Fundamental Considerations in Language Testing. Oxford: Oxford University Press, 1990.

BACHMAN, L.F./PALMER, A. S. Language Testing in Practice: Designing and Developing Useful Language Tests. Oxford: Oxford University Press, 1996.

JIA, Y. Ethical Standards for Language Testing Professionals: An Introduction to Five Major Codes, 2009. Dostupné z <http://jalt.org/test/jia1.htm>

KANE, M.T. An argument-based approach to validity. Psychological Bulletin 112, s. 527-535. 1992.

McNAMARA, T. Language testing. Oxford: Oxford University Press, 1992.

NEWFIELDS, T. Voices in the field: An interview with Randy Thrasher. SHIKEN: The JALT Testing & Evaluation SIG Newsletter, 5 (1), s. 12-14. 2001. Dostupné z http://jalt.org/test/thr_new.htm

SPOLSKY, B. Measured words: The development of objective language testing. Oxford: Oxford University Press, 1995.

AERA/APA/NCME. Standardy pro pedagogické a psychologické testování. (Klimusová, H. Trans.). Testcentrum: Praha, 2001 (Original work published in 1999).

WEIR, C.J. Language Testing and Validation: An evidence-based approach. Palgrave MacMillan, 2005.

ÚDAJE O AUTOROVI:

Mgr. Martina Hulešová, M.A.
 Výzkumné a testovací centrum
 Ústav jazykové a odborné přípravy Univerzity Karlovy
 Vratislavova 10
 128 00 Praha 2
 Kontakt: martina.hulesova@ujop.cuni.cz