

K PROBLEMATICE VALIDITY TESTŮ STUDIJNÍCH PŘEDPOKLADŮ – zkušenosti s jejich tvorbou na FF UP v Olomouci

LUCIE VIKTOROVÁ & MIROSLAV CHARVÁT

Katedra psychologie, Filozofická fakulta, Univerzita Palackého v Olomouci

Abstrakt: *Součástí přijímacího řízení na vysoké školy jsou často i tzv. testy studijních předpokladů, se kterými se pojí řada psychometrických otázek. V příspěvku se zaměřujeme na problematiku validity těchto testů, a to jak po obsahové, konstruktové, tak i po prediktivní stránce. Z českých i zahraničních výzkumů a zkušeností vyplývá zejména nejasnost definice tzv. „předpokladů ke studiu“ i toho, co je možné považovat za „akademický úspěch“. Zmiňujeme zde výsledky některých studií na toto téma i praktické problémy, které mohou při zjišťování validity testů studijních předpokladů vyvstat. V našich úvahách se přitom opíráme o vlastní zkušenosti z analýz Testu předpokladů ke studiu na FF UP (SPF).*

Klíčová slova: *testy studijních předpokladů, konstruktová validita, prediktivní validita, přijímací řízení na VŠ*

Afiliace ke grantu: *Studentská grantová soutěž FF UP IGA FF_2013_11*

Úvodem

V dřívějších dobách byl přístup k vysokoškolskému vzdělání v první řadě otázkou financí, sociálního postavení či kádrových posudků a politické loajality, tento model se ovšem postupem času transformuje (Urbášek, 2008). Pozornost je přitom čím dál více přenášena na schopnosti, které musí uchazeč o studium na vysoké škole doložit. Většinou mezi nimi nalezneme prokázání znalostí a zájmu o obor, předložení seznamu prostudované relevantní literatury či prací souvisejících s předmětem studia, případně potvrzení o praxi v dané oblasti.¹ K těmto spíše oborově specifickým požadavkům se i u nás od roku 1996, kdy společnost Scio přišla na trh s testem obecných studijních předpokladů (OSP), stále častěji připojuje i ověřování jakési všeobecné schopnosti

¹ V souvislosti s přijímacím řízením na FF UPOL jsme s trochou nadsázky konstatovali, že vedle těchto „manifestních“ požadavků je důležitým nárokem na uchazeče schopnost dorazit včas, na správné místo, se všemi doklady, umět se dle příjmení dostavit ke správné komisi/stolku, pozorně si přečíst zadání a vyplnit všechny požadované kolonky, včetně např. vlastního jména.

úspěšně studovat na vysoké škole. A právě tzv. předpoklady ke studiu a jejich testování se stávají častým námětem diskuzí nejen mezi potenciálními uchazeči, ale také odbornou pedagogickou veřejností (např. Atkinson, 2001; Frey, & Detterman, 2003; Ruijssenaars, Castelijns, & Hamers, 1992). S konstrukcí testů studijních předpokladů se totiž pojí několik základních psychometrických otázek – zvláště, pokud si jednotlivé vysoké školy či fakulty tvoří tyto testy samy.²

V prvé řadě jde o jejich validitu, a to jak obsahovou, tak i kritériovou a prediktivní. Jednotná definice „předpokladů ke studiu“ stejně jako „akademického úspěchu“ totiž navzdory logickému úsudku prozatím neexistuje (Atkinson, 2001). Někteří výzkumníci v této oblasti (např. Atkinson, 2001; Frey, & Detterman, 2003) odkazují na úmysl testů studijních předpokladů měřit jakousi schopnost uvažování (*reasoning ability*), jejíž odlišnost od obecné inteligence je ovšem nejasná. Jiní (např. Kuncel et al. 2001, 2010) poukazují na konkrétní složky, které se testy nejčastěji pokoušejí zachytit – verbální myšlení, kvantitativní operace, analytické uvažování, řešení problémů apod. – a určitou definici kruhem, tj. že testy studijních předpokladů měří schopnost uchazeče úspěšně studovat na (dané) vysoké škole.³ Podobně je možné akademický úspěch operacionalizovat jednak jako dosažení titulu, ale také jako dobu potřebnou k jeho dosažení, celkový průměr známek za studium, průměr známek v prvním ročníku, dosažená stipendia, počet publikovaných článků nebo schopnost získat dobře placené zaměstnání či vysokou vedoucí pozici (Burton, & Ramist, 2001; College Entrance Examination Board, 2002; Kuncel, Hewlett, & Ones, 2001). A to vše za situace, kdy samotné univerzity či fakulty mnohdy nemají jasně stanoveny, o jaké uchazeče konkrétně mají zájem (National Association for College Admission Counselling, 2008). Otázka, o čem tedy výsledek v testech studijních předpokladů vlastně vypovídá, proto raději bývá v praxi často decentně obejit.

Při naší analýze Testů předpokladů ke studiu na FF UP (SPF) jsme se s těmito problémy v praxi potýkali i my, a proto jsme se rozhodli věnovat tento příspěvek několika úvahám a obtížím, které pro nás v průběhu prvních fází naší práce vyvstaly. Zaměříme se tedy blíže na validitu testů studijních předpokladů a s tím související výzkumy české i zahraniční.

Konstrukt a obsah „předpokladů ke studiu“

Počátky dnešních testů studijních předpokladů lze vystopovat do roku 1926, kdy byl v Americe jako součást přijímacího řízení na některé univerzity poprvé uveden tzv. Scholastic Aptitude Test (SAT) (Lawrence, Rigol, Van Essen, & Jackson, 2002). Od té

² Rozsáhlou praxi má v tomto směru Masarykova univerzita v Brně, která vlastní testy studijních předpokladů (TSP) vytváří a na sedmi z devíti fakult používá již deset let – viz např. Šavelková (2013, 29. dubna). Právě Masarykovou univerzitou se při vývoji SPF inspirovala i FF UP.

³ Tento popis najdeme např. i na stránkách Masarykovy univerzity (<http://www.muni.cz/tsp>) a Univerzity Palackého v Olomouci (<http://www.ff.upol.cz/skupiny/zajemcum-o-studium/bakalarske-a-magisterske-studium/prijimaci-rizeni/#c17455>).

doby byl tento test několikrát přepracováván, vznikaly jeho konkurenční verze (např. ACT – American College Testing, GRE – Graduate Record Examination pro přijímání do magisterských studií nebo TSA – Thinking Skills Assessment ve Velké Británii) a posléze posloužil jako podklad pro podobné testy v neanglicky mluvících zemích. Zároveň s tímto vývojem plynula i diskuze o tom, co má či nemá být náplní testů studijních předpokladů, které vlastnosti, schopnosti a dovednosti by se měly u uchazečů o studium na vysoké škole zjišťovat a jak lze vlastně definovat a měřit „úspěch ve studiu“. Pojďme si nyní projít jednotlivé body o něco podrobněji.

První verze SAT obsahovaly 7 verbálních a 2 matematické subtesty, přičemž jejich obsah byl postupně upravován (z verbálního oddílu např. byly vypuštěny definice a antonyma, naopak ponechány byly analogie, doplňování vět a porozumění textu). V současnosti sestává SAT z oddílů kritického čtení, matematiky a sepsání krátké eseje (Lawrence, Rigol, Van Essen, & Jackson, 2002). Pro srovnání: test ACT ověřuje znalosti a dovednosti v oblastech „anglický jazyk“, „matematika“, „čtení“, „vědecké uvažování“ a volitelné „eseji“, TSA je zaměřeno na „řešení problémů“ (numerických a prostorových), „kritické myšlení“ a „písemnou úlohu“ a GRE obsahuje subtest verbální (zaměřený na porozumění textu, kritické uvažování a užití slov), kvantitativní (s matematickými úlohami) a „analytické psaní“ se dvěma úlohami. Podobnou rozmanitost najdeme i u testů studijních předpokladů v českém prostředí – například Scio rozšířilo svá OSP (sestavující z verbálního, analytického a kvantitativního myšlení) o subtesty „kritické myšlení“ a „abstraktní uvažování“.

Ani tyto specifické schopnosti nejsou v praxi zdaleka tak jednoznačně oddělitelné a měřitelné, jak se může zdát. Autoři podílející se na tvorbě přijímacích testů na Masarykově Univerzitě v Brně (tyto TSP se nepoužívají na FSS MU) se například zaměřili na individuální styly řešení položek, které se vyskytují v subtestu prostorové představivosti. Jejich výsledky hovoří o tom, že celkový skóre subtestu prostorové představivosti může minimálně z části odrážet nejen vizuálně-analogovou schopnost prostorové představivosti, ale také verbálně-analytickou schopnost, která v určitém typu položek může při řešení prostorovou představivost víceméně nahradit. Tento fakt je pak podle nich nutno vzít v úvahu při interpretaci výsledku takového testu (Jelínek, Květoň, & Vobořil, 2013). Tato drobná ukázka podrobné analýzy fungování položek jednoho ze subtestů ukazuje, jak složitá může otázka studijních předpokladů jako celku být. Už z pohledu na pouze částečně se překrývající kategorie různých testů studijních předpokladů ovšem musí nutně vyvstat otázka, co tedy ony „studijní předpoklady“ vlastně jsou a které části by měly obsahovat, nebo jinak: Jaké je nejvhodnější měřítko uchazečovy schopnosti úspěšně studovat na vysoké škole?

Je zajímavé, že na rozdíl od anglosaských zemí u nás většinou nenajdeme oddíl odpovídající „písemné úloze“, resp. eseji, a naopak se často setkáváme s oddílem „všeobecný/kulturní přehled“. Může jít o kulturně podmíněnou tradici, kdy v českých školách není psaní esejí zakořeněno natolik jako např. v USA či Velké Británii, a naopak

je u nás kladen větší důraz na určité penzum faktických znalostí.⁴ Potíž je, že právě určení, které znalosti jsou natolik „základní“ či „nutné“, že je musí každý uchazeč o vysokoškolské studium bezpodmínečně znát, je často arbitrární a závislé na autorech testu, popř. jejich oborovém zaměření. (Např. důležitost znalosti data bitvy u Slavkova může jinak hodnotit historik, jinak lingvista a zcela jinak psycholog.) To samo o sobě zpochybňuje obsahovou validitu takového testu a navíc může mnohdy přispívat k argumentu, že testy studijních předpokladů znevýhodňují starší uchazeče, kteří nejsou „čerstvě po maturitě“, případně nejsou absolventy gymnázií, kde se předpokládá „širší všeobecně-znalostní“ základ. Zároveň však mohou být znevýhodněni i mladší uchazeči, kteří se již neučí vše nazpaměť, nýbrž projektově, s možností využít internet a jiné databáze, a nesoustředí se tolik na hromadění informací v hlavě, ale spíše na procesy zpracování informací.

V souvislosti s nejnovějšími revizemi SAT byl v zahraničí diskutován rozdíl mezi ověřováním „dosazeného úspěchu/výkonu“ (*achievement*, měřící úroveň znalostí a opírající se o středoškolské kurikulum, tzn. co byl student schopen si odnést ze střední školy) a „schopností/potenciálu“ (*aptitude*, snažící se zachytit méně obsahově závislé a více přenositelné kompetence, tzn., co vše by mohl zvládat v budoucnu) a jejich důležitosti při přijímacím řízení na vysokou školu (např. Atkinson, 2001; Zwick, 2007). Jedním z hlavních argumentů, proč upustit od důrazu na všeobecné studijní předpoklady ve prospěch oborově specifických testů, přitom byla právě provázanost s výukou na střední škole, tj. možnost uplatnit již nabyté vědomosti při postupu do další úrovně studia namísto „procvičování, jak zvládnout testy“ (Atkinson, 2001). Na druhou stranu fakt, jak dobře dovedou uchazeči v testu skórovat, pakliže mají možnost se na něj připravit, může být brán jako demonstrace jejich potenciálu k učení (Ruijsenaars, Castelijns, & Hamers, 1992), čímž by tyto testy svůj úkol splnily. Příprava na test ze strany uchazečů je přitom i z hlediska např. české legislativy, která ustanovuje povinnost zveřejňovat již proběhlá znění testů, nevyhnutelným fenoménem, s nímž musí všichni tvůrci testů počítat, i kdyby jejich primárním cílem bylo zjistit spíše schopnosti „nenacvičené“. K této problematice se vyjadřuje např. National Association for College Admission Counselling (2008).

Zůstává však otázka, nakolik se poté „studijní předpoklady“ liší od jiných konstruktů, například od faktoru obecné inteligence. Ve srovnávacích analýzách bývá jako zástupce IQ testů zjišťujících obecnou inteligenci nejčastěji používán Ravenův test progresivních matic; mezi jinými zjistili např. Frey a Detterman (2003) na souboru 104 studentů bakalářského programu signifikantní korelaci mezi skórem v Ravenových progresivních maticích a skórem testu SAT ($r = 0.483$, $p < 0.001$) (po korekci omezeného výběru čili zohlednění vysoké homogenity testované skupiny, tzv. *restriction of range*, korelace stoupla na $r = 0.72$). Podobně společnost Scio (Scio, 2012, 9. května) analyzovala 290 účastníků testu OSP, kteří zároveň vyplnili IQ test předkládaný společností Mensa ČR

⁴ Poněkud smutnější interpretace by byla, že eseje představují příliš časově zatěžující a na vyhodnocení náročný úkol, kterému se české vysoké školství raději vyhýbá.

(blíže neidentifikovaný, leč formátem velmi podobný Ravenovým progresivním matricím) a zjistila korelaci $r = 0.5$.⁵ Ačkoliv tyto údaje každá ze skupin interpretuje poněkud jinak⁶, faktem zůstává, že schopnosti a vlastnosti měřené pomocí testů studijních předpokladů souvisejí určitým způsobem s obecnou inteligencí – a koneckonců i s výsledky oborových testů, jak odkazuje například Zwick (2007).⁷

Kritéria a prediktivní validita

Tím se postupně dostáváme k nutnosti v první řadě specifikovat schopnosti, dovednosti a vlastnosti, které jednotlivé vysoké školy či fakulty od svých potenciálních studentů vyžadují, ne vždy však explicitně uvádějí. Přitom se nemusí jednat pouze o nabyté vědomosti či matematicko-analytické a verbální dovednosti, ale také o osobní a sociální kompetence (např. schopnost pracovat samostatně i v týmu, pod stresem či za využití dostupných zdrojů, schopnost empatie), charakterové vlastnosti (jako je dodržování morálních a etických zásad či přítomnost vůdčích schopností) anebo vlastnosti související s motivací (např. vytrvalost, pečlivost, dodržování termínů apod.) (srov. College Entrance Examination Board, 2002). Pokud řekneme, že přítomnost těchto charakteristik u uchazeče je předpokladem ke studiu na (dané) vysoké škole, pak by se logicky od takto specifikovaných požadavků mělo odvíjet složení přijímacích testů, resp. náplň celého přijímacího řízení. Problémem v praxi ovšem je, že v mnoha případech (snad až na talentové zkoušky) se testování (rozumějme zjišťování pomocí psychodiagnostických metod) jakýchkoliv jiných vlastností osobnosti, než je IQ, jeví jako obtížně prosaditelné. Důvodem pro tento stav mohou být obavy z napadení regulérnosti přijímacího řízení či osočení z diskriminace. Často je tento výběr implicitně obsažen až v rámci ústních pohovorů, kde se ovšem členové přijímacích komisí daných oborů mohou spoléhat jen na svůj kvalifikovaný odhad a úsudek. Když tuto situaci srovnáme s výběrovými řízeními na důležité posty v soukromém sektoru, musíme konstatovat, že obavy z použití jiných než výkonových testů v rámci přijetí na VŠ jsou možná zbytečné.

Zároveň bychom jako tvůrci či „stanovitelé“ těchto požadavků měli mít na mysli, za jakým účelem chceme, aby měl uchazeč/ budoucí student právě takové vlastnosti, schopnosti a dovednosti. Je našim cílem vybrat „úspěšného studenta“, nebo „úspěšného absolventa“, tj. člověka, který se dovede v budoucnosti dobře uplatnit ve své zvolené profesi? V odborné literatuře se s tímto tématem setkáme při diskuzi o operacionalizaci akademického úspěchu: Stejně jako u požadavků kladených na potenciální studenty se i zde jedná o široké spektrum kritérií – od známkového průměru v prvním roce studia či celkového průměru známek za studium (tzv. *grade point average*, GPA) přes dosažení

⁵ V článku o této studii však chybí některé důležité metodologické údaje nutné pro kvalitní interpretaci výsledků.

⁶ „SAT je adekvátní mírou obecné inteligence“ (Frey, & Detterman, 2003, s. 7) vs. „[...]závislost mezi výsledkem testu IQ a průměrným percentilem v OSP není příliš silná [...]. Jako nejpravděpodobnější se jeví vliv různého zaměření obou testů.“ (Scio, 2012, 9. května, odst. 12 a 13).

⁷ Tím se i u testů studijních předpokladů dostáváme k otázce férovosti vzhledem ke kulturnímu původu, socioekonomickému statutu a jazykové vybavenosti uchazečů, která byla v minulosti široce řešena právě v souvislosti s IQ testy; její diskuze by ovšem dalece přesáhla formát tohoto příspěvku.

bakalářského titulu a nástup do magisterského studia až po úspěch v budoucím profesním životě, získání významného vůdčího postavení, akademických ocenění aj. (srov. Rigol, 2003).

Ve výzkumné praxi je coby kritérium predikční validity testů studijních předpokladů nejčastěji používán výše zmíněný známkový průměr (jak v prvním roce studia, tak i za studium celkově), výsledky jednotlivých analýz se ovšem mnohdy různí. Burton a Ramist (2001) sestavili přehled studií od roku 1980 zkoumajících SAT a známky ze střední školy jako prediktory známkového průměru v prvním roce a průměru za celou dobu studia. Průměrná korelace mezi skórem SAT a GPA v těchto studiích činí $r = 0.36$ (s rozpětím 0.22–0.52), pro SAT a známky v prvním roce studia potom $r = 0.35$. Průměrná korelace mezi známkami ze střední školy a GPA přitom činí $r = 0.42$ (s rozpětím 0.34–0.57), pro známky ze střední školy a průměr v prvním roce studia $r = 0.36$. Kombinací středoškolských známek a skóru SAT se korelace s GPA pohybuje kolem $r = 0.52$, korelace se známkami v prvním roce studia poté okolo $r = 0.44$.⁸ Je tedy vidět, že známky ze střední školy souvisí s výsledky na VŠ minimálně stejně silně jako skór v SAT; za této situace je možné se ptát, zda je z hlediska efektivity či inkrementální validity skutečně nutné testy studijních předpokladů administrovat. Z českých výzkumů na toto téma lze uvést např. studii Zváry a Anděla (2001) na studentech MFF UK po prvním a druhém roce studia. Vedle výsledku testu studijních předpokladů byl mezi prediktory úspěšného absolvování prvního a druhého ročníku (tj. dosažení minimální kreditové hranice pro postup do dalšího ročníku) v logistické regresi např. průměr známek ze střední školy, počet bodů z lepšího ze dvou oborových testů a fakt, zda byl uchazeč přijat s prominutím přijímací zkoušky. Poslední proměnná přitom dosahovala stabilně (napříč všemi obory MFF) nejvyšších beta-koeficientů, následována průměrem známek ze střední školy a poté přibližně do stejné míry body z oborového testu a testu studijních předpokladů. Autoři však došli k závěru, že vzhledem k velkým směrodatným odchylkám „úspěšnost studia jen poměrně volně souvisí s veličinami zaznamenávanými v přijímacím řízení a že i celý model logistické regrese může být příliš zjednodušený“ (Zvára, & Anděl, 2001, s. 339).

Pokud jde o další prediktory a kritéria akademického úspěchu, Höschl a Kožený (1997) provedli analýzu na 92 studentech 3. LF UK v prvních třech letech studia, přičemž jako prediktory pro známkové průměry v jednotlivých ročnících a souhrnně za 3 roky studia byly použity výsledky oborových testů (matematika, fyzika, chemie), průměr známek ze střední školy (z předmětů matematika, fyzika, český jazyk), bodové hodnocení vstupního pohovoru (ověřujícího motivaci, sociální zralost uchazeče a jeho schopnost reprodukce textu) a skóre dvou škál Trojdimenzionálního osobnostního dotazníku (škála sentimentality a attachmentu). Kromě různě silných interkorelací mezi jednotlivými proměnnými z výsledků mnohonásobných lineárních regresí vyplynulo, že nejlepšími prediktory celkového známkového průměru za 3 roky studia jsou výsledky oborového

⁸ Hodnoty korelačních koeficientů se po korekci omezeného výběru a různého systému známkování zvednou o 0.04–0.07 bodu pro GPA a o 0.17–0.19 pro známky v prvním roce studia; celkově se tedy pohybují okolo $r = 0.41$ – 0.64 .

testu z fyziky a průměr známek z fyziky ze střední školy, které společně vysvětlovaly 21 % rozptylu, doplněné o hodnocení studentovy motivace a výsledek škály attachmentu (dohromady vysvětlující 32 % rozptylu). Podobnou strukturu a hodnoty vykazovala i lineární regrese pro známkový průměr v prvním roce studia, pouze hodnocení motivace bylo nahrazeno schopností reprodukce textu. Ve druhém a třetím roce studia se již složení proměnných měnilo více a i celkový podíl vysvětlené variance klesal. V praxi to tedy znamená, že z výsledků oborových testů a známek z oborově relevantních předmětů ze střední školy lze do určité míry usuzovat na známkový průměr na VŠ, rozhodně se ale nejedná o kompletní výčet proměnných, které toto kritérium „úspěšného studia“ ovlivňují.

Co se dalších možností definic „akademického úspěchu“ týče, Kuncel, Hezlett a Ones (2001) provedli metaanalýzu na datech studentů navazujících magisterských programů, přičemž využili jejich známkového průměru za bakalářské studium a výsledku testů GRE (verbálního, kvantitativního, analytického a oborového) k odhadu jednak známkového průměru v prvním ročníku a celkově za navazující magisterské studium, jednak např. fakultního hodnocení, počtu vydaných článků či konferenčních příspěvků a doby potřebné k dosažení magisterského titulu. Výsledné korelace (korigované pro omezení výběru a další statistická zkreslení) mezi subtesty GRE a GPA, průměrem v prvním ročníku a fakultním hodnocením se opět pohybují v rozmezí od $r = 0.32-0.50$, vyšších hodnot ovšem dosahují korelace s oborovými než s ostatními „obecnými“ subtesty. Pokud jde o souvislost s dobou získání titulu, počtem vydaných příspěvků a jejich citací, jsou korelace s GRE všeobecně slabší ($r = 0.02-0.28$). Zdá se, že při hodnocení predikční validity testů studijních předpokladů tedy poměrně záleží na tom, co stanovíme kritériem akademického úspěchu.

Autoři Sackett, Borneman, & Connelly (2008) publikovali rozsáhlé zhodnocení meta-analytických studií s velkými vzorky s cílem prověřit často zmiňované výhrady a kritiky namířené na schopnosti testů studijních předpokladů SAT vzhledem k jejich využitelnosti pro oblast výběru zaměstnanců a přijímacího řízení na vysoké školy. Došli k závěru, že (a) testy studijních předpokladů jsou obecně validní pro jejich zamýšlené použití, tj. predikují celou řadu krátkodobých a dlouhodobých aspektů akademického úspěchu i pracovního výkonu, (b) validita testů není artefaktem socioekonomického statusu, (c) koučování a nácvik zvládání těchto testů není hlavním určujícím faktorem úspěšnosti v testech a že (d) testy obecně nevykazují zkreslení u členů národnostních menšin.

Obdobně povzbudivé jsou i výsledky zprávy zaměřené na posouzení validity SAT po zavedení určitých změn ve složení a obsahu jednotlivých subtestů, která proběhla v roce 2005. Tato zpráva dokumentuje výsledky založené na velkém vzorku 196 364 studentů přijímaných na 110 čtyřletých vysokých škol a univerzit na podzim 2006 a jejich výsledků po dokončení prvního ročníku vysoké školy v květnu/červnu 2007. Výsledky ukazují, že změny provedené v SAT nezměnily nic na jejich prediktivní validitě týkající se výkonu v prvním roce studia na vysoké škole. Ve všech sledovaných institucích, se

nedávno přidaná část psaní eseje ukázala jako nejsilněji predikující ze všech tří jednotlivých sekcí SAT. Dále se zjistilo, že jako nejsilnější pro predikci akademického úspěchu v prvním roce (FYGPA) se jeví kombinace prediktorů: průměru známek na střední škole (HSGPA) a celkového SAT skóre (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008).

Další potíže s výpočtem predikční validity

Kromě výše uvedených „teoretických úvah“ o volbě vhodných kritérií se při zjišťování prediktivní validity testů předpokladů ke studiu můžeme setkat ještě s ryze praktickými obtížemi. Patrně tou nejvýznamnější je již zmíněné „omezení výběru“ na uchazeče s vysokými skóry – ti jediné jsou totiž ke studiu přijati a tudíž pouze od nich jsou dostupná data jako známkový průměr za první a další ročníky studia.⁹ Toto snížení variability skóru způsobuje i snížení maximální možné zjistitelné korelace mezi proměnnými (Urbánek, Denglerová, & Širůček, 2011). Jak jsme měli možnost pozorovat na výše uvedených výzkumech, je tento problém možné do určité míry statisticky korigovat, což však nic nemění na faktu, že „potenciál k úspěšnému studiu“ u nepřijatých uchazečů zůstává neověřen.

Na tomto místě bychom mohli diskutovat, zda potenciál nepřijatých uchazečů není natolik nízký, že se o „akademickém úspěchu“ nedá hovořit ani ve smyslu přijetí ke studiu, to bychom ovšem museli zanedbat problémy s obsahovou a konstruktovou validitou testů studijních předpokladů, jinak bychom se snadno mohli dostat do objasňování pomocí tautologie. Zároveň je třeba mít na paměti, že bodová hranice přijetí je taktéž často stanovena arbitrárně a stejně jako u jiných psychodiagnostických nástrojů ani v tomto případě není možné určit, z jakého důvodu by měl být rozdíl byt' jen jediného bodu na této hranici klinicky signifikantní.

Z procesního hlediska navíc může být pro výzkumníka problém se k datům od jednotlivých uchazečů dostat. Při naší práci na analýzách Testů předpokladů ke studiu na FF UP jsme například řešili otázku, zda se data uvedená v dokumentech spojených s přijímacím řízením dají považovat za majetek univerzity a tudíž je možné je využít k výzkumu bez nutnosti explicitního písemného souhlasu každého z uchazečů. Ne všechna data se také musí nutně nacházet v jediné databázi (či vůbec v elektronické podobě, jako např. známky z maturitního vysvědčení, dokládané mnohdy až u zápisu ověřenou kopií dokumentu¹⁰) a jejich spárování při zachování anonymity jedince a bezpečnosti dat může být poměrně značným logistickým oříškem. Pomineme-li byrokratickou zátěž spojenou s celým procesem, „třešničkou na dortu“ při zjišťování prediktivní validity např. celofakultně jsou poté uchazeči, kteří se hlásí na více různých

⁹ Samozřejmě ne všichni přijatí uchazeči navíc studium dokončí; chceme-li tedy k predikci kompletní data, je mortalita ve vzorku ještě vyšší.

¹⁰ Právě z důvodu obtížného ověření pravdivosti údajů se také některá data, která by byla pro výpočet prediktivní validity vhodná (např. známky z posledních dvou ročníků střední školy) vůbec nemusí od uchazečů vyžadovat.

oborů (příčemž stejný skóre v testu může u různých oborů vést k různým výsledkům ve smyslu „přijat/nepřijat“), případně pak na dvouobory s různou váhou skóre testu studijních předpokladů v celkovém výpočtu bodů pro jednotlivé obory.

Kde není reliabilita, není ani validita

Tato základní poučka nás nutí alespoň stručně zmínit i otázky spojené s určováním reliability testů studijních předpokladů. Ta je dalším problémovým bodem při zajišťování validity, a to především v situaci, kdy z finančních, časových, personálních či jiných důvodů není možné zajistit plnohodnotné pilotní testování. Zanedbatelným problémem není ani fakt, že konkrétní znění položek testů je nutné chránit před prozrazením. Test-retestová reliabilita tedy v tomto případě prakticky nepřipadá v úvahu. S tím souvisí i problematika položkové analýzy, která by v případě pretestu umožnila vyloučit z „ostré“ verze špatně diferencující či jinak nekvalitní položky. Pokud však máme k dispozici jen analýzu „post hoc“, můžeme se pouze pokusit dodatečně identifikovat typ položek, které by se (ne)měly objevovat v příštích ročnících testu. Stanovení reliability paralelních verzí (většinou se v přijímacím řízení používá mnoho variant) může rovněž probíhat spíše ex post a spíše zprostředkovaně pomocí posuzování podobnosti výsledků z jednotlivých verzí. Při porovnávání vnitřních konzistencí též můžeme narazit například na fakt, že ne všechny verze testů musí vykazovat stejné hodnoty koeficientů reliability.

Jedním ze slibných přístupů, jak se tohoto úkolu spojeného s reliabilitou zhostit, je využití teorie odpovědi na položku (item response theory). Pokud bychom se však chtěli opírat o tento způsob analýzy, může nám činit potíže požadavek unidimenzionality testu (Jelínek, Květoň, & Vobořil, 2011). Jedná se o úvahu, že odpověď na danou položku je ovlivněna pouze jedním latentním rysem (v našem případě onou „schopností úspěšně studovat na vysoké škole“). V momentě, kdy se rozhodneme tento předpoklad aplikovat na celý test (jenž se většinou skládá z různých oddílů jako je kritické či verbální myšlení, prostorová představivost apod.), tím de facto zpochybníme nutnost testy na jednotlivé oddíly členit, čímž se dostaneme zpět k otázce validity, tj. co vlastně tyto testy ve skutečnosti zjišťují.

Diskuze a závěry

Ověřit v praxi konstruktovou a predikční validitu testů studijních předpokladů tedy není nic jednoduchého. V zahraničí, zejména pak v USA, je této problematice věnována značná pozornost již řadu desetiletí. U nás se zájem o tuto problematiku začal objevovat až po demokratizaci země, kdy objektivita přijímacího procesu začala být brána mnohem více v potaz. V současné době se u nás pravděpodobně i pod vlivem dlouholetého a z hlediska validity i poměrně uspokojivého výzkumu testů studijních předpokladů v USA rozvíjí právě tato forma písemných přijímacích testů (většinou ovšem bez psaní eseje). Vše je založeno na jednoduché úvaze, že pokud budeme vytvářet

testy podobné svou strukturou i obsahem testům SAT, bude to též fungovat. Tento postup je ekonomický a asi i racionální, nicméně validita neplatí univerzálně, nemůžeme tedy na její prověřování zcela rezignovat.

Tento proces ovšem podle našich zkušeností z tvorby a analýzy testů SPF na FF UP v Olomouci komplikuje celá řada překážek. Zjednodušeně řečeno, se potýkáme zejména s faktem, že testy studijních předpokladů nemají ustálenou podobu, jejich obsah a některé parametry jsou už z podstaty věci proměnlivé. Navíc jsou tyto testy až do chvíle jejich jednorázové aplikace chápány jako přísně tajné, načež musí být ze zákona po skončení přijímacího řízení zveřejněny. Vysoké školy jsou tak pod tlakem veřejnosti na férovost a objektivitu těchto testů, v důsledku čehož nejsou psychometrické vlastnosti těchto testů většinou bohužel předmětem veřejné akademické diskuse. Jakákoliv významná změna v přijímacím procesu musí navíc projít interním schvalovacím procesem a musí také být včas uveřejněna, což celý proces zkvalitňování testů dosti prodlužuje. Jak z toho ven? Odpovědi nejsou jednoduché a hledáme je jen obtížně. Roli hraje mnoho faktorů.

V našem případě jsme se rozhodli, že prvním krokem by mělo být definování toho, co od uchazečů očekáváme, jaké jejich kvality hledáme. To se samo o sobě ukázalo jako téma pro samostatnou výzkumnou substudii. Potom teprve by měla nastupovat otázka, jak lze tato specifika zjišťovat, nicméně z důvodů zvyklosti, ekonomičnosti a také bezpochyby i jisté setrvačnosti je forma omezena pouze na jednoduše vyhodnitelný písemný test. V tomto kontextu je důležité zmínit, že přijímací proces je samozřejmě často složený z více částí a z více kol. Je tedy vhodné si připustit, že písemné testy studijních předpokladů často slouží jako jakýsi prvotní filtr, který redukuje velké množství uchazečů na zvládnutelné počty, u kterých pak můžeme zjišťovat více podrobností například v ústním pohovoru. U prvotního filtru tedy musíme vybírat spíše na základě těch fundamentálních a univerzálních vlastností, kam určitě nepatří specifické oborové znalosti, ale spíše dílčí kognitivní funkce či některé fasety inteligence jako je například kritické, logické či analytické myšlení.

V dalším kroku si vysoká škola či fakulta může rozhodnout, zda tento problém delegovat na jinou instituci a svým způsobem ho outsourcovat nebo ho zajišťovat vlastními zdroji. Přenesení odpovědnosti na někoho jiného nemusí nutně vadit, nemusí to být nikterak fatální a celý přijímací proces může i tak docela dobře fungovat, tj. může být validní a objektivní. Může to ale být nebezpečné v tom ohledu, že nebudeme svou zakázku jasně formulovat, nakonec nemusíme dostat takové studenty, jaké bychom si představovali a jaké požadujeme. Hrozí zde i jistá spíše nežádoucí komercializace celého přijímacího aktu. V druhém případě má vysoká škola či fakulta možnost mít celý proces od začátku do konce pod kontrolou, ale za cenu jistě vynaložené energie. Musí si především zajistit tým odborníků, kteří se na tuto problematiku dlouhodobě zaměří. Musí se také pečlivě věnovat složitým přípravám všech detailů organizačního procesu. Nelze to dělat tzv. „na koleně“ nebo jen jako nutné zlo. Jedině dlouhodobá práce, skládající se z pečlivé přípravy testů, pokud možno bezchybné administrace a vyhodnocení, následné detailní

psychometrické analýzy výsledků, interpretace chování testů v minulých ročnících a dlouhodobého sledování kritérií úspěšnosti ve studiu, může přinést kýženou kvalitu do přijímacího procesu.

Domníváme se též, že vzhledem ke specifickým požadavkům na studenty u jednotlivých oborů nelze do budoucna vždy spoléhat pouze na tuto písemnou testovou formu přijímacího řízení, byť by se jí podařilo dovést k větší dokonalosti. Doufáme, že i v českém kontextu bude přibývat článků zaměřených na toto téma a že bude možné se více vzájemně inspirovat.

Literatura

- Atkinson, R.C. (2001). Achievement Versus Aptitude Tests in College Admissions. *Issues in Science and Technology*, 18(2), 1-9. Získáno z http://works.bepress.com/richard_atkinson/28/.
- Burton, N.W., & Ramist, L. (2001). *Predicting Success in College: SAT Studies of Classes Graduating Since 1980*. College Entrance Examination Board. Získáno z <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2001-2-predicting-college-success-sat-studies.pdf>.
- College Entrance Examination Board. (2002). *Best practices in admissions decisions: A report on the third College board conference on admission models*. New York: College Board Publications. Získáno z <http://research.collegeboard.org/sites/default/files/publications/2012/7/misc2002-1-best-practices-admissions-decisions.pdf>.
- Frey, M. C., & Detterman, D. K. (2003). Scholastic Assessment or *g*? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. *Psychological Science*, 15(6), 373–378. Získáno z <http://www.psychologicalscience.org/pdf/ps/frey.pdf>.
- Höschl, C., & Kožený, J. (1997). Predicting academic performance of medical students: the first three years. *American Journal of Psychiatry*, 154, 87–92.
- Jelínek, M., Květoň, P., & Vobořil, D. (2011). *Testování v psychologii: Teorie odpovědi na položku a počítačové adaptivní testování*. Praha: Grada.
- Jelínek, M., Květoň, P., & Vobořil, D. (2013). Skryté aspekty v testování prostorové představivosti: identifikace uplatňovaných stylů řešení položek. *Československá Psychologie*, 57(4), 297-306.
- Kobrin, J.L., Patterson, B.F., Shaw, E.J., Mattern, K.D., & Barbuti, S.M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average*. New York: The College Board.
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions. *Current Directions in Psychological Science*, 19, 339-345. Získáno z http://psychoserver.psych.arizona.edu/JJBAReprints/PSYC621/Kuncel_Hezlett_Current%20Directions%20in%20Psychological%20Science-2010%20%28Cog%20Ability%20Testing%20Hiring%20and%20Admissions%29.pdf.

- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162–181. Získáno z <http://internal.psychology.illinois.edu/~nkuncel/gre%20meta.pdf>.
- Lawrence, I., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2002). *Research Report No. 2002-7: A Historical Perspective on the SAT: 1926–2001*. College Entrance Examination Board. Získáno z <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2002-7-historical-perspective-sat-1926-2001.pdf>.
- National Association for College Admission Counselling (2008). *Report of the Commission on the Use of Standardized Tests in Undergraduate Admission*. Získáno z http://www.nacacnet.org/research/PublicationsResources/Marketplace/Documents/TestingComission_FinalReport.pdf.
- Rigol, G. W. (2003). *Admissions decision-making models. How U.S. institutions of higher education select undergraduate students*. College Entrance Examination Board. Získáno z http://www.collegeboard.com/prod_downloads/press/adm_decision_making.pdf.
- Ruijsenaars, A.J.J.M., Castelijns, J.H. M., & Hamers, J.H.M. (1992). The validity of learning potential tests. In J. H. M. Hamers, A. J. J. M. Ruijsenaars, & K. Sijtsma (Eds). *Learning Potential Assessment: Theoretical, Methodological and Practical Issues* (69–82). UK: Taylor & Francis.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-Stakes Testing in Higher Education and Employment. *American Psychologist*, 63(4), 215–227. doi:10.1037/0003-066X.63.4.215
- Scio. (2012, 9. května). *Srovnávací analýza OSP a IQ*. Získáno z: <http://www.scio.cz/vyzkum/analyzy/osp-iq.asp>.
- Šavelková, J. (2013, 29. dubna). *Deset let testu studijních předpokladů*. Získáno z: <http://www.online.muni.cz/udalosti/3612-deset-let-testu-studijnich-predpokladu#.U10kKlB7I9p>.
- Urbánek, T., Denglerová, D., Širůček, J. (2011). *Psychometrika: měření v psychologii*. Praha: Portál.
- Urbášek, P. (2008). *Vysokoškolský systém v letech tzv. normalizace*. Olomouc: Univerzita Palackého v Olomouci.
- Zvára, K., & Anděl, J. (2001). Souvislost výsledků přijímacího řízení s úspěšností studia na MFF. *Pokroky matematiky, fyziky a astronomie*, 46(4), 304–312.
- Zwick, R. (2007). *College admission testing*. National Association for College Admission Counseling. Získáno z <http://www.nacacnet.org/research/PublicationsResources/Marketplace/Documents/TestingWhitePaper.pdf>.

Podpořeno z projektů OPVK:

SOVA-21 – Internacionalizace, inovace, praxe: sociálně-vědní vzdělávání pro 21. století,

CZ.1.07/2.2.00/28.0225

INZA – Inovací bakalářských studijních programů k lepší zaměstnatelnosti,

CZ.1.07/2.2.00/28.0238



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ