

Univerzita Palackého v Olomouci

Filozofická fakulta

Katedra psychologie

**RELIABILITA A VALIDITA SPF
(TEST PŘEDPOKLADŮ KE STUDIU NA FF UP)**

RELIABILITY AND VALIDITY OF THE SPF
(LEARNING POTENTIAL TEST FOR FF UP)



Magisterská diplomová práce

Autor: Lucie Viktorová
Vedoucí práce: Mgr. Miroslav Charvát, Ph.D.
Studijní program: Psychologie
Obor: Psychologie (PS)

Olomouc

2014

Prohlášení

Místopřísežně prohlašuji, že jsem magisterskou diplomovou práci na téma: „Reliabilita a validita SPF (Test předpokladů ke studiu na FF UP)“ vypracovala samostatně pod odborným dohledem vedoucího diplomové práce a uvedl/a jsem všechny použité podklady a literaturu.

Vdne

Podpis

Poděkování

Na tomto místě bych chtěla poděkovat všem členům týmu IGA FF_2013_011, zvláště pak svému vedoucímu, Mgr. Miroslavu Charvátovi, PhD., za vzájemnou duševní i faktickou podporu při tvorbě této práce. Jmenovitě patří velký dík také PhDr. L. Vobořilovi, PhD. za zprostředkování komunikace se studijním oddělením FF UP a CVT, Ing. P. Grigárkové jakož i dalším studijním referentkám za laskavé umožnění přístupu k potřebným datům a RNDr. J. Roglovi za ochotnou součinnost při generování výpisů z databází.

Obsah

Úvod.....	5
Teoretická část diplomové práce	6
1 Vysvětlení základních pojmů.....	6
2 Přístup k vysokoškolskému vzdělání – současná praxe a potíže	9
3 Co je akademický úspěch?.....	11
4 Prediktory akademického úspěchu a požadavky kladené na uchazeče o studium na VŠ	15
4.1 „Studijní předpoklady“ – co to je a jak je zjistit?	18
4.1.1 Konstrukt „studijních předpokladů“	21
4.2 Dosavadní výkon vs. výkonový potenciál a „naučitelnost“ na testy	24
4.3 Prediktivní validita testů studijních předpokladů a dalších faktorů.....	25
4.4 Doporučení pro vysoké školy k ověřování předpokladů ke studiu.....	32
5 Test předpokladů ke studiu na FF UP (SPF)	38
6 Kritéria kvality testových metod.....	41
6.1 Standardizace a objektivita	41
6.2 Férovost	43
6.3 Reliabilita.....	45
6.3.1 Reliabilita paralelních forem	47
6.3.2 Reliabilita vnitřní konzistence	48
6.4 Validita.....	52
6.4.1 Obsahové zdroje důkazů o validitě.....	52
6.4.2 Empirické/Kritériové zdroje důkazů o validitě.....	55
6.4.3 Konstruktové zdroje důkazů o validitě	58
6.5 Průnik mezi reliabilitou a validitou testu: Položková analýza.....	59
Výzkumná část diplomové práce	62

7	Výzkumný problém, cíle práce a hypotézy	62
8	Popis zvoleného metodologického rámce a metod	64
8.1	Typ výzkumu a metody získávání dat	64
8.2	Etické problémy a způsob jejich řešení	65
8.3	Metody zpracování a analýzy dat	66
8.3.1	Posouzení reliability testů SPF 2011-2013	66
8.3.2	Obsahová validita a položková analýza	68
8.3.3	Kritériová prediktivní a inkrementální validita	69
8.4	Výzkumný soubor	72
9	Výsledky	75
9.1	Výsledky posouzení reliability testů SPF 2011-2013	75
9.2	Výsledky zhodnocení obsahové validity a položkové analýzy SPF	79
9.3	Výsledky ověřování prediktivní a inkrementální validity testů SPF	83
10	Diskuze	90
11	Závěry	99
	Souhrn	101
	Seznam použitých zdrojů a literatury	104

Abstrakt diplomové práce

Abstract of thesis

Přílohy diplomové práce

Úvod

V diskuzi o podobě terciárního vzdělávání je jedním z aktuálních problémů kvalitní výběr budoucích studentů z uchazečů o studium na vysoké škole. Vzhledem k velikosti poptávky po vysokoškolském studiu není zcela možné podstoupit s každým uchazečem osobní pohovor, proto školy hledají nějaké časově, finančně i personálně úsporné „prvotní síto“, které by umožnilo odlišit „potenciálně úspěšné“ studenty od těch „méně schopných“. Filozofická fakulta Univerzity Palackého v Olomouci (dále jen FF UP) si coby toto síto počínaje akad. rokem 2011/2012 zvolila vlastní Test předpokladů ke studiu na FF UP (SPF), který je povinnou součástí přijímacího řízení na všechny bakalářské obory Filozofické fakulty.

A právě s tímto testem a jeho rolí v přijímacím řízení se pojí projekt IGA FF_2013_011: *Testy studijních předpokladů jako součást přijímacího řízení na vysokou školu a jejich potenciál k predikci akademického úspěchu*. Jeho cílem je v první řadě ověřit psychometrické vlastnosti, tj. především reliabilitu a validitu testu SPF, ale v kontextu odpovědi na otázku, zda test plní svůj účel, také mj. zformulovat profil žádoucího uchazeče o studium na FF UP, který by měl být v Testech předpokladů ke studiu (SPF) zohledněn.

Tato diplomová práce vznikla jako součást projektu FF_2013_011 a jejím cílem je popis reliability a validity testů SPF z let 2011, 2012 a 2013. Zároveň se zabývá i problematikou definice „akademického úspěchu“, „studijních předpokladů“ a konstrukcí položek a testů zamýšlených k jejich ověřování, stejně jako některými aspekty přijímacího řízení na vysoké školy obecně. Doufáme, že naše zjištění pomohou tvůrcům testů i politiky přijímacího řízení na FF UP ve zkvalitňování tohoto procesu v budoucích letech.

Teoretická část diplomové práce

1 Vysvětlení základních pojmů

S ohledem na psychometrickou povahu této práce považujeme za vhodné nejprve objasnit několik zásadních termínů, s nimiž budeme na následujících stránkách pracovat. Některé koncepty (např. prediktivní a kritériová validita, reliabilita atp.) budou později rozvedeny v samostatných kapitolách, pro čtenáře by však mělo být příjemnější, pokud se seznámí s těmito pojmy dříve, než se pustí do čtení dalšího textu. Za základní termíny pro naše účely považujeme následující:

Reliabilita. Reliabilita patří mezi základní kritéria kvality testových, potažmo psychodiagnostických metod, a je v podstatě podmínkou jejich validity. Urbánek, Denglerová a Širůček (2011, s. 94) uvádějí dvě obecné definice reliability, a to:

„1. ... charakteristika psychodiagnostické metody, která uvádí relativní nepřítomnost chyb v měření,

2. ... jiné označení pro spolehlivost nebo přesnost metody měření“.

K jejím odhadům můžeme použít různých přístupů (klasická testová teorie, teorie odpovědi na položku aj.) a metod (stabilita skóru testovaného v čase, paralelní formy testů, vnitřní konzistence testů apod.), hovořit proto budeme i o různých *druzích* reliability. Podstatné je v tento moment ještě upozornit, že celý koncept reliability se zakládá na systematické *variabilitě* měřených skórů (resp. znaků), tzn. že v testovaném souboru je co do měřeného znaku potřeba určitá heterogenita, jinak se reliabilitu nepodaří ověřit či bude zkreslená (nižší) než v případě méně homogenního vzorku. Tomuto problému se říká *omezení variability* (angl. *restriction of range*) a může nastat především tehdy, když je vzorek vybírán právě na základě hodnot znaku, u něhož zjišťujeme reliabilitu způsobu jeho měření (Urbánek, Denglerová, & Širůček, 2011). To by mohl být i případ našich analýz testů SPF, neboť uchazeči o studium na vysoké škole mohou již předběžně vykazovat vyšší studijní předpoklady než zbytek populace. Omezení variability pak může také zasáhnout výpočty prediktivní validity testů SPF vzhledem ke známkovému prospěchu studentů v prvních letech studia, neboť tento výběrový soubor bude ještě užší a bude se skládat pouze z uchazečů, kteří (patrně) dosáhli v testu SPF (či ještě v oborových testech) nejvíce bodů – tedy opět z více homogenní skupiny. Důvod, proč na tento problém upozorňujeme již zde, je ten, že termín „omezení variability“ se objevuje i v některých studiích, jejichž výsledky chceme diskutovat ještě před hlubší analýzou pojmu „reliabilita“.

Validita. Podobně jako reliabilita patří i validita mezi kritéria kvality psychodiagnostických a testových metod. Do češtiny se nejčastěji překládá jako „platnost“, tzn. zda test skutečně měří to, co jsme jím měřit chtěli (Urbánek, Denglerová, & Širůček, 2011). Říčan (1977, s. 70) navíc připomíná důležitou „schopnost testu diagnostikovat, predikovat, měřit něco víc než „sebe sama“. Naráží tím na účel veškerého testování v psychologii, totiž snahu na základě pozorovatelného chování, výkonu či sebehodnocení usuzovat na latentní proměnné (někdy nazývané také *faktory*) – tj. nejčastěji nějaké schopnosti, vlastnosti, dovednosti nebo znalosti posuzovaného. V tomto kontextu se pak i při posuzování validity používaných nástrojů setkáme s následujícími pojmy: *Konstrukt* je obecné označení pro měřené charakteristiky (vlastnosti, schopnosti, dovednosti atp., viz níže), které má odrážet jejich de facto spekulativní podstatu; jak poznamenávají Urbánek, Denglerová a Širůček (2011, s. 79) nejsou konstrukty „v podstatě ničím jiným než víceméně přijatelnými pojmovými nástroji pro vymezení a občas i vysvětlení určitých pozorování“. Jde tedy o to, že veškeré koncepty jako „inteligence“, „studijní předpoklady“, ale i „temperament“ nebo „osobnost“ nejsou ve skutečnosti něčím viditelným či hmatatelným a coby psychologové se spíše domýšlíme, že něco takového existuje (na základě určitých pravidelností v pozorování). Proto mohou být některé konstrukty velmi nejasně definované, což pak podstatně komplikuje jejich měření či ověřování; s tímto problémem se ostatně budeme potýkat napříč naší prací. *Kritériem* bývá v průběhu zjišťování validity měrného nástroje většinou myšlena závislá proměnná, tzn. nějaký jev, který se pomocí dané metody snažíme předpovědět – např. jaký budoucí známkový průměr půjde očekávat u různých uchazečů o studium na VŠ. *Prediktory* jsou poté veškeré proměnné, které se k této předpovědi pokusíme využít, jako třeba známky ze SŠ či maturitního vysvědčení (Camara, 2004a). V naší práci se někdy bude zdát, že se tyto termíny, resp. proměnné slévají; např. může účast v dobrovolnických organizacích být jednak prediktorem, jednak jedním z kritérií akademického úspěchu. Zároveň může být i slovo „kritérium“ použito obecně pro označení nějaké sledované či potenciální proměnné, aniž by to nutně znamenalo proměnnou závislou. Pokusíme se tyto situace rozlišit, ačkoliv ani v primární literatuře není tato hranice vždy zcela jasná.

Schopnosti, kognitivní schopnosti. V Hartlově (2004, s. 244) psychologickém slovníku je schopnost definovaná jako „soubor předpokladů nutných k úspěšnému vykonávání nějaké činnosti nebo dovednosti; schopnosti se vyvíjejí na základě vloh učení“. Carroll (1993) ovšem upozorňuje na v literatuře (a to jak laické, tak odborné) mnohem širší pojetí

schopnosti (angl. *ability*) coby vyjádření potenciálu (angl. *aptitude*) „něco umět/moci“. Táže se tedy, jak obecné či specifické mohou schopnosti být, zda je možné považovat je za *vlastnost* (trvalý rys) osobnosti a do jaké míry jsou schopnosti kvantifikovatelné. Všechny tyto otázky totiž ovlivňují výzkum a praxi v oblasti vzdělávání, především pak odvětví zabývající se testováním výkonu. Jak uvidíme dále v textu, diskuze nad tím, zda lze „schopnosti“ či „potenciál“ zlepšovat učením (tzn. každý člověk má šanci své schopnosti zdokonalovat) nebo zda se jedná o něco vrozeného (a tudíž by bylo „nefér“ vybírat např. uchazeče o vysokoškolské studium právě jen a pouze na základě takového měření), totiž do značné míry ovlivnila vývoj v oblasti přijímacích řízení na vysoké školy. *Kognitivní schopnosti* je poté možno nahlížet jako specifickou podkategorii, někdy nazývanou též *intelektové schopnosti* (Hartl, 2004) – i když i tato zaměnitelnost může způsobovat názorové rozkoly mezi odborníky ve vzdělávání (viz kapitola o konstruktivním všeobecných studijních předpokladů). Většinou v sobě zahrnují procesy spojené se zpracováním mentálních informací (Carroll, 1993), tzn. v nejširším pojetí i paměť, pozornost, myšlení (vč. jednotlivých myšlenkových operací jako analýza, syntéza, indukce, dedukce, usuzování apod.) a imaginaci. V kontextu testování ve vzdělávání se však jednotliví odborníci často omezují pouze na určitý výsek těchto procesů, např. pouze na myšlení či paměť. V této práci budeme pojmy „schopnosti“, příp. „kognitivní schopnosti“ používat v jejich nejširším smyslu, budeme ovšem průběžně diskutovat jejich významové nuance. Zároveň je budeme považovat za odlišné od *dovednosti* čili vyloženě „učení získané dispozice ke správnému, rychlému a úspornému vykonávání určité činnosti vhodnou metodou“ (Hartl, 2004, s. 51) a také od *znalosti* či *vědomosti*, přičemž poslední dva termíny budeme používat zaměnitelně; oba se vztahují k zapamatovaným informacím a jejich propojování (Hartl, 2004). Termín *charakteristiky* pak budeme používat souhrnně pro schopnosti, dovednosti, vlastnosti a znalosti/vědomosti případně pro kombinace těchto pojmů.

Uvedený výčet termínů rozhodně není vyčerpávající, měl by však umožňovat i méně zkušenému čtenáři základní orientaci v psychometrických aspektech dalších kapitol. Pakliže se v textu objeví další potenciálně méně známé či komplikovanější pojmy, budou již vysvětleny přímo ad hoc.

2 Přístup k vysokoškolskému vzdělání – současná praxe a potíže

Navzdory současné krizi v zaměstnatelnosti absolventů některých oborů je vysokoškolské vzdělání ve společnosti obecně vnímáno ve spojení s vyšší prestiží, lepšími vyhlídkami na budoucí pracovní uplatnění a především jeho platové podmínky, a tedy jako poměrně žádoucí (Bollinger, 2004; Urbášek, 2008). Kapacita vysokých škol ovšem není neomezená a je proto nezbytné, aby jejich zdroje byly využity co nejefektivněji – tj. pokud možno ke kvalitní výuce studentů, kteří mají odpovídající schopnosti, dovednosti a motivaci zvolený obor dostudovat a profitovat ze svého studia v pracovní praxi. Tím vstupuje do popředí problematika výběru budoucích studentů z uchazečů o studium na vysoké škole.

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). Prudký, Pabian a Šima (2010) hovoří o přechodu od *elitního* k *univerzálnímu* vzdělávání. Elitní vzdělávání se vyznačuje především tím, že zasahuje pouze <15% aktuální věkové kohorty, která by o studium na vysoké škole mohla jevit zájem. Pokud se počet uchazečů, resp. studujících zvýší ke 30-50%, nastává fáze *masového* terciárního vzdělávání, a po překročení této hranice je možné vzdělávání označit za *univerzální*. Autoři zde vycházejí z konceptu M. Trowa (1997, in Prudký, Pabian, & Šima, 2010), který dále popsal deset oblastí, v nichž se aktuální podoba vysokoškolského vzdělávání odráží; mezi tyto oblasti patří např. vnímání přístupu k terciárnímu vzdělání, funkce vysoké školy, podoba kurikula a způsobů výuky, rozhodovací mechanismy na akademické půdě, standardy kvality a kritéria pro vstup na vysokou školu. Česká republika se při tom dle odhadů autorů (co do podílu studujících z relevantní věkové kohorty, tzn. 18-19-letých) v současnosti nachází ve fázi univerzálního vzdělávání, která se vyznačuje vnímáním vysokoškolského studia víceméně jako povinnosti – zvláště ve vyšších společenských vrstvách – a důrazem na volný přístup k VŠ vzdělání, pakliže má uchazeč motivaci studovat.

Této charakteristice poněkud odporuje běžná praxe mnoha českých vysokých škol (typická pro elitní, či v určité podobě i pro masové vzdělávání) vybírat si ze svých uchazečů budoucí studenty. Kopíruje tím poměrně rozšířený model ze zahraničí, především pak USA, kde jsou prestižní univerzity vnímány jako výzkumné instituce (ne

nepodobné jiným firmám), které mají zájem na maximálním přínosu studenta pro univerzitu a obráceně (College Entrance Examination Board, 2002). A právě v tomto bodě se dostáváme k zásadnímu problému, totiž nakolik se vlastně univerzita či její fakulty v ČR mohou chovat jako „jiní zaměstnavatelé“ a stanovovat si základní požadavky na schopnosti, dovednosti, vlastnosti a znalosti svých uchazečů, aniž by tyto byly označeny za „znevýhodňující“ nějakou skupinu obyvatel a důvodem pro přijetí odvolání uchazeče. Laird (2004) upozorňuje na podstatný rozdíl právě oproti USA, totiž že ve Spojených státech patří mezi základní akademické svobody právo fakulty na volbu těch, které bude učit. V českém Zákoně č. 111/1998 Sb. o vysokých školách podobnou formulaci nenajdeme: §4 – Akademické svobody a akademická práva v tomto směru hovoří pouze o svobodě volby studijního zaměření, potažmo svobodě výuky a otevřenosti různým názorům. Naopak článek 33, odst. 1 Listiny základních práv a svobod (Zákon č. 2/1993 Sb.) udává, že „každý má právo na vzdělání“. Ačkoliv je možný výklad tohoto odstavce pouze v kontextu základního, případně středoškolského vzdělání, je celkem snadné si představit jeho hodnotu coby argumentu v odvolání pro přijetí na VŠ.

Můžeme si tedy vůbec z právního hlediska dovolit stanovovat požadavky na uchazeče? Pokud ne, jak je možné, že u nás vůbec přijímací řízení na VŠ probíhá? A pokud ano, proč bychom, podobně jako zaměstnavatelé, nemohli trvat na (praxí opodstatněných) specifických požadavcích – nejen znalostních, ale třeba i osobnostních – bez jejichž splnění by uchazeč „neprošel“? A to aniž bychom byli v kterémkoliv bodě označeni za „diskriminační“? Zodpovězení těchto otázek považujeme za základní krok potřebný ke zkvalitnění procesu výběru budoucích studentů FF UP.

Pokud jde o samotné charakteristiky, které jsou u uchazečů o studium na vysoké škole vyhledávány a hodnoceny, jejich typ i rozsah často souvisí s definicí pojmu „akademický úspěch“. Stejně jako u mnoha dalších konceptů ve vzdělávání a psychologii se však jedná o definici napříč odborníky velmi různorodou a komplexní, a proto se jí budeme podrobněji věnovat v následující kapitole.

3 Co je akademický úspěch?

V zahraničních (ale i některých českých) výzkumech k prediktivní/kritériové validitě testů studijních předpokladů je akademický úspěch nejčastěji operacionalizován jako známkový průměr za celé studium (angl. *grade point average*, dále jen GPA) či v prvním roce studia (angl. *first year grade point average*, dále jen FYGPA). Existuje ovšem několik závažných argumentů, že tato definice není dostačující či dokonce vhodná:

V prvé řadě Laird (2004) poukazuje na skutečnost, že známkování na vysoké škole je mnohdy přísnější na technických a přírodovědných oborech než na oborech humanitních a společenských (o rozdílech ve známkování napříč jednotlivými předměty a vyučujícími nemluvě). Tento fakt může způsobovat tzv. „efekt stropu“, totiž „hromadění“ známek na jednom konci spektra, a tudíž velmi nízkou variabilitu v kritériu. Ta pak obratem může zkreslit predikční schopnost testů vzhledem ke známám v prvním a dalších ročnících VŠ; jedná se o určitou manifestaci dříve zmíněného omezení variability (viz také kap.6.4.2).

Zadruhé může použití FYGPA a GPA v českých podmínkách vykazovat větší zkreslení co do variability známkování napříč předměty než v zahraničí, konkrétně pak v USA, odkud pochází nejvíce výzkumů na toto téma. Ve Spojených státech jsou totiž studenti bakalářského studia nejprve zapsáni na *univerzitu* a absolvují velkou část základních kurzů společně, než si zvolí svou *oborovou* specializaci. Tím by se teoreticky mělo zmenšovat zkreslení způsobené výše zmíněnou rozdílnou praxí ve známkování napříč různými vědními odvětvími (a s nimi spojenými programy), případně předměty. I tak ovšem Sackett, Borneman a Connelly (2008) upozorňují, že toto zkreslení celkových známek způsobené výběrem různých kurzů (s různou náročností) v USA existuje a je dobré jej při výpočtech prediktivní validity testů ošetřit, jsou-li k tomu dostupná data – například známky z jednotlivých předmětů. Žoudlík (2009) uvádí, že nejlepší operacionalizací GPA by byl výpočet průměrů známek jen z relevantních předmětů, resp. předmětů, kterými prošli všichni studenti; v případě našich analýz testů SPF na FF UP ovšem tento postup bohužel aplikovat nelze, neboť takto podrobná data jsou nám nedostupná.

Zatřetí – a co je asi nejdůležitější – je akademický úspěch ve své podstatě multidimenzionální, přičemž FYGPA a GPA se soustředí pouze na jednu složku, totiž kognitivní výkon či schopnosti. Mohu sice odrážet motivaci, úsilí či svědomitost, ale opět:

vzhledem k různým způsobům známkování napříč jednotlivými obory, předměty a vyučujícími nejsou spolehlivým ukazatelem těchto faktorů (Camara, 2004a). Pokud zůstaneme (převážně) na akademické půdě, je možné úspěch operacionalizovat také jako dosažení titulu či dobu potřebnou k jeho dosažení, aktivitu v hodinách, získaná stipendia a další ocenění, počet publikovaných článků nebo schopnost následně získat dobře placené zaměstnání či vysokou vedoucí pozici (Burton, & Ramist, 2001; College Entrance Examination Board, 2002; Kuncel, Hezlett, & Ones, 2001).

Ani tento výčet ovšem není zcela kompletní. Willingham (1985, in Camara, 2004a) provedl metaanalýzu studií zabývajících se definicí akademického úspěchu a identifikoval společná kritéria, která zmiňovali zástupci vysokých škol i studentů, totiž intelektuální a osobní růst, etičnost, komunikační dovednosti, schopnost vést a schopnost analyzovat problémy a osvojovat si nové dovednosti. Mezi další hodnotitelné faktory by mohlo patřit např. využití zdrojů fakulty či přispění společnosti dobrovolnickou a jinou činností. Podobně Schmitt, Oswald a Gillespie (2004) podotýkají, že univerzity často argumentují profitem studentů z jejich zdrojů na několika úrovních (znalosti, dovednosti, ale také osobnostní a sociální rozvoj), tyto oblasti však zřídka bývají zařazeny jako kritéria akademického úspěchu ve validizačních studiích metod výběru uchazečů v přijímacím řízení. Navrhují proto tato kritéria rozšířit o následující oblasti, jež odpovídají vzdělávacím cílům, jak je definují na svých stránkách vybrané zahraniční vysoké školy: faktické znalosti, zvědavost a zájem o kontinuální učení; vytrvalost; představa budoucí kariéry; společenská zodpovědnost a angažovanost; etičnost a integrita; multikulturní tolerance; interpersonální a komunikační dovednosti a schopnost vést; adaptabilita a životní dovednosti; fyzická a duševní kondice a (d)ocenění hodnoty umění. Tomuto výčtu také odpovídá model úspěšného pregraduálního/bakalářského studia, jak jej sestavil Camara (2004a) a který zahrnuje:

a) Úspěch v samostatném získávání informací, popř. samostatném výzkumu.

Úspěšný student by měl tedy prokázat svou schopnost efektivního využívání technologií a informačních zdrojů, přicházet s novými nápady a způsoby řešení jakož i aplikovat ověřené výzkumné a analytické metody ve své vlastní praxi. Tento bod v sobě zahrnuje i předpoklad kritického čtení a naslouchání a (expertní) zpracování deklarativních vědomostí. Jeho indikátorem by mohla být např. kvalita průběžných seminárních a závěrečných bakalářských prací studentů.

- b) Rozvoj komunikačních a interpersonálních schopností a dovedností.** Student by se tedy měl být schopen srozumitelně vyjadřovat slovem i písmem a řídit se pravidly mezilidské komunikace. Zároveň by měl ukázat schopnost tolerance, týmové spolupráce, ale případného vedení (leadershipu). V ideálním případě by se tedy měl zapojit do studentských či dobrovolných spolků a organizací a podílet se na jejich aktivitách.
- c) Akademické výkony a self-management.** Zde je myšleno především dosažení přiměřeného známkového průměru (průběžně i na konci studia – přičemž tento konec by v ideálním případě neměl přesáhnout standardní dobu studia) a dalších studijních úspěchů (např. účast na konferencích, publikace nebo různá ocenění). Předpokladem je projevené úsilí a motivace – svědomitost, docházka a aktivní účast na hodinách, případně absolvování přednášek a kurzů nad rámec povinné výuky.
- d) Plán budoucí kariéry.** Úspěšný student by měl být cílevědomý, orientovaný na vlastní rozvoj, a v průběhu bakalářského studia si ujasnit své další profesní směřování – ať již v následujícím magisterském studiu či přímo v pracovní praxi.

Vyhledávané vlastnosti budoucích **absolventů** lze tedy zhruba rozdělit do několika oblastí (dle College Entrance Examination Board, 2002):

- a) Vědomosti a intelektové schopnosti.** Jedná se např. o schopnost kritického a analytického myšlení, jak za pomoci kvantitativních, tak kvalitativních metod; schopnost kreativního řešení problémů a přenosu vědomostí a dovedností napříč různými oblastmi; odpovídající ústní a písemný projev a v neposlední řadě zájem o vědecké bádání a připravenost k celoživotnímu učení.
- b) Integrativní schopnosti.** Sem lze zařadit pochopení provázanosti vědy, technologie a společnosti, jejich morálních hodnot, a schopnost tyto poznatky aplikovat ve svém profesním i osobním rozhodování.
- c) Sociální a kulturní kompetence.** Do této oblasti spadá schopnost pracovat jak samostatně, tak i v týmu; schopnost navazovat přátelské vztahy a řešit konflikty; znalost občanských práv a povinností a angažovanost v globálních i místních komunitních aktivitách.

d) Osobní kompetence. Mezi osobní kompetence lze zařadit schopnost efektivně vyhledávat zdroje k uspokojení osobních i profesních potřeb, schopnost práce se stresem a udržování zdravého životního stylu a fyzické jakožto i duševní kondice.

Z tohoto výčtu je už na první pohled zřejmé, jak nesnadné může být tato kritéria posuzovat, popřípadě vůbec kvantifikovat či měřit, a to u tisícovek studentů, kteří každoročně na vysokých školách odpromují. Přesto existují určité možnosti a konkrétní nástroje, které lze využít, přičemž některé z nich zmíníme i dále v této práci. Za důležité považujeme na tomto místě zdůraznit, že je-li cílem univerzity nabrat skupinu uchazečů s různorodými schopnostmi, zájmy či talentem, pak by setrvání u jediného kritéria způsobilo, že veškeré „nové“ prediktory budou hodnoceny pouze na základě inkrementální validity vůči tomuto jedinému kritériu. Ta přitom nemusí být příliš vysoká, neboť např. známky mohou být celkem dobře pokryty právě testy studijních předpokladů nebo známkovými průměry ze střední školy. Mohlo by tedy dojít k „zavržení“ prediktorů, které jsou potenciálně velmi přínosné k předpovědi jiných kritérií úspěchu – která zároveň také opomeneme brát v úvahu (Camara, 2004a).

Kromě toho si lze všimnout, že mnohá kritéria akademického úspěchu by mohla stejně tak dobře být považována za vstupní znalosti, schopnosti a dovednosti potřebné ke studiu na VŠ (čili jeho prediktory) - a při pohledu na výčty charakteristik požadovaných od uchazečů na vysoké školy zjistíme, že tomu tak mnohdy skutečně je. Důvodem, proč se „obě strany rovnice“ často slévají, je jednak fakt, že vlastnosti žádoucí u uchazečů jsou dále využívány během terciárního studia a jejich rozvoj tedy může být vnímán jako „akademický úspěch“, jednak mohou tyto v přijímacím řízení sloužit jako nejlepší prediktor „sebe sama“ (např. Camara, 2004b).

Tím se tedy dostáváme k nutnosti v první řadě specifikovat schopnosti, dovednosti a vlastnosti, které vysoké školy od svých potenciálních studentů vyžadují (ne vždy však také explicitně uvádějí). Jednotlivé vysoké školy či fakulty přitom mohou klást na výše zmíněné složky různý důraz, který bude (v ideálním případě) patrně odrážet deklarované poslání a hodnoty dané VŠ. Tato problematika je opět komplexnější, proto se jí budeme věnovat v následující kapitole.

4 Prediktory akademického úspěchu a požadavky kladené na uchazeče o studium na VŠ

Conley (2003) v publikaci „Standards for Success“ zveřejňuje výsledky dvouletého výzkumu, během něhož byli dotázáni členové 400 fakult z celkem 20 amerických univerzit, co by měli umět uchazeči o studium na VŠ, aby úspěšně zvládli základní předměty bakalářského programu.¹ Více než na specifické oborové znalosti kladli dotázaní důraz na „zvyky“ studentů: kromě kognitivních dovedností (kritické a analytické myšlení a řešení problémů, schopnost jasného vyjádření se slovem i písmem atp.) jmenovali například zvědavost, schopnost posoudit důvěryhodnost různých informačních zdrojů a pracovat s nimi nebo ochotu akceptovat konstruktivní kritiku a vyrovnat se s případnými neúspěchy či náročnými úkoly. „Oborové znalosti“ pak byly rozřazeny do šesti okruhů (anglický jazyk, resp. jazykové znalosti spojené se čtením, psaním, hledáním informací a kritickým myšlením v mateřském jazyce; cizí jazyk, matematika, přírodní vědy, společenské vědy a umění) a tvoří jakýsi soubor doporučení pro střední školy, jaké znalosti, schopnosti a dovednosti by ve svých studentech měly rozvíjet. Přitom „úspěch“ je těmito standardy definován jako „schopnost dostatečně dobře prospívat v základních bakalářských předmětech, splnit obecné vzdělávací požadavky a pokračovat v oborové specializaci“ (Conley, 2003, 9).

Campbell, McCloy, Oppler a Sager (1993, in Camara, 2004a) hovoří o 3 hlavních dimenzích přispívajících k výkonu ve škole: *deklarativní vědomosti, procedurální znalosti a dovednosti a motivace*. Kromě toho jmenují ještě osm výkonových komponent, které mohou v určitých situacích sehrát roli moderátorů: specifické znalosti ve vztahu k zadanému úkolu, obecné rozumové schopnosti, komunikační dovednosti, úsilí, disciplína, týmová spolupráce, supervize a vůdčí schopnosti.

S ohledem na tyto informace je možné se tázat, jací tedy *jsou* současní vysokoškoláci. Podrobný rozbor této problematiky by vydal na samostatnou publikaci, uvedme ale alespoň postřehy Pavlíčkové (nedat.) na základě jejích zkušeností z výuky, konzultací a přijímacích pohovorů se studenty na Vysoké škole zemědělské v Brně. Konstatuje, že většina současných studentů je spíše submisivních, pasivních a nepřilíží zvyklých o sobě přemýšlet (z hlediska svých vlastností). Studenti mají dle autorky velmi

¹ Je třeba mít na paměti, že americký vysokoškolský systém funguje jinak než český: Studenti jsou nejprve „obecně“ zapsáni na danou univerzitu, kde musí projít základními společnými předměty, a teprve poté si volí vlastní specializaci.

špatné komunikační dovednosti, zvláště pak verbální vyjadřování a schopnost argumentace, přičemž často nedovedou dodržet daný časový limit a neskákat jinému mluvčímu do řeči. Disponují zároveň vesměs poměrně dobrou pamětí a lehce nadprůměrnou inteligencí, přičemž se jim lépe daří v nonverbálních testech. Autorka zmiňuje také pozitivní korelaci mezi měřeným IQ u přijímacích zkoušek a prospěchem v prvním ročníku této školy, podrobnosti ovšem neuvádí. Současně sděluje, že tvořivost studentů je poměrně nevyrovnaná co do individuální a týmové tvořivosti (s tendencí jednotlivců tíhnout k jednomu či druhému extrému), přičemž složka originality a fluence je u nich spíše průměrná až podprůměrná². Jakkoliv subjektivní mohou tyto postřehy být, lze si z nich vytvořit určitý obraz o současných vysokoškolácích (minimálně na VŠZ) a porovnat je s vlastnostmi, které bychom ideálně u studentů vysokých škol chtěli.

Jelikož výběr vhodných faktorů i vzhledem k mnohoznačné definici akademického úspěchu je bez konkrétního zakotvení velmi obtížný, nabízí se pro potřeby vysokých škol postup uplatňovaný při náboru zaměstnanců v pracovním prostředí. Při něm je třeba nejprve popsat náplň, znaky a výkonové dimenze jednotlivé pracovní pozice, z nich pak odvodit možné schopnosti a dovednosti potřebné k jejímu úspěšnému zastávání a poté ke každému faktoru navrhnout či vyvinout několik možných způsobů jeho ověřování či měření. Mezi tyto způsoby mohou patřit hodnotící škály (sebeposuzovací či pro pověřené zaměstnance), psychologické osobnostní či výkonové testy, archivní záznamy nebo řešení modelových situací (Camara, 2004a).

Na FF UP by se v podstatě jednalo o aplikaci principu doporučeného Lairdem (2004), totiž v první řadě pečlivě formulovat cíle a hodnoty Univerzity Palackého a Filozofické fakulty, a z nich následně odvozovat požadavky na uchazeče, které hledá. V rámci naší rešerše webových stránek UP a FF UP jsme zatím takto ucelenou (obecnou) formulaci nenalezli; stránka *O univerzitě* (Univerzita Palackého v Olomouci, 2013) nabízí pouze stručné shrnutí základních dat o UP, FF UP pak pod odkazem *O fakultě* hovoří o principu svobodné volby studia a osobního přístupu ke studujícím, a zmiňuje podporu kreativního myšlení a studentských aktivit jakož i „důraz na začlenění studentů do odborných a kulturních činností“ (Filozofická fakulta Univerzity Palackého v Olomouci, 2013, odst. 3).

² Zajímavá je také konstatace, že zjišťování tvořivosti studentů již při přijímacích zkouškách se ukázalo vzhledem k poptávce po studiu jako organizačně nezvladatelné (opět bez uvedení dalších podrobností).

Na tomto místě je třeba podotknout, že máme na mysli „misi a vizi“ UP či FF UP jako celé instituce. Pro jednotlivé obory na FF UP lze dohledat jejich charakteristiky a plánované profily absolventů, případně i požadavky pro přijímací zkoušky (PZk), nevyhne se však několika potížím: Zaprvé jsou tyto popisy u různých oborů různě (více či méně) obsáhlé, přičemž nutno říct, že se většinou drží na úrovni obecných formulací a příliš nezabíhají do konkrétností. Zvláště pak v požadavcích k PZk často nalezneme pouze odkaz na testy studijních předpokladů a oborové testy či ústní pohovor. Zadruhé se jedná o popis *oboru* a profilu *absolventa*, tj. studenta, který daným programem projde, z nichž nejsou explicitně odvozeny žádné konkrétní požadavky na vstupní schopnosti, vlastnosti či dovednosti *uchazeče*. Zatřetí se, vzhledem k rozmanitosti oborů nabízených na FF UP, lze tázat, zda vůbec existuje nějaký „prototyp studenta“ či jeho charakteristik, které by byly stejně důležité či uplatnitelné bez ohledu na studovaný obor. Touto otázkou se zabývá bakalářská práce M. Tošenovské (2014), která zkoumá profil ideálního uchazeče a studenta tak, jak jej vnímají vyučující některých oborů na FF UP. Z jejích rozhovorů ovšem vyplývá, že i někteří katoři mají potíže určit, jaké jsou žádoucí vlastnosti, schopnosti či dovednosti jejich „nejlepších“ studentů (přičemž obsah slova „nejlepší“ si vyučující volí sami) a co bychom tedy měli u uchazečů ověřovat, natož pak jakým způsobem. Zároveň diskutuje problematiku, zda opravdu chceme „uniformní“ studenty, nebo zda je naopak pro FF UP přínosem rozmanitost osobností i v rámci, nejen napříč jednotlivými obory, která by samozřejmě byla hůře postihnutelná jediným testem ověřujícím úzké spektrum kognitivních dovedností.

Pokud bychom sáhli do dalších zdrojů, nabízí se rozhovor s novým rektorem UP, prof. Jaroslavem Millerem (Rychlík, 29. ledna 2014), v němž předkládá svou vizi UP coby mezinárodní univerzity, kde bude anglický (příp. i německý) jazyk využíván na stejné úrovni jako český a kde bude vyžadováno dodržování morálních a etických principů. Vystává tedy otázka, zda už u přijímacího řízení do bakalářského studia trvat např. na (ověření) znalosti anglického jazyka. Mnoho uchazečů, kteří na střední škole studovali jiný cizí jazyk než angličtinu, totiž může namítnout, že takový požadavek je vůči nim diskriminační.

Dle Lairda (2004) je v tomto kontextu zásadní rozhodnutí, zda by přijímací řízení mělo vybírat ty, kteří až doposud dosahovali výjimečných výsledků, nebo ty, kteří by mohli dosahovat výjimečných výsledků právě na dané univerzitě – a co tyto „výjimečné výsledky“ jsou. Camara (2004b) například říká, že informace o dosavadních úspěších a

aktivitách uchazečů mohou být vzhledem k predikci budoucích aktivit a úspěchů hodnotné, neboť lidské chování, postoje a hodnoty jsou poměrně konzistentní. Mimoškolní aktivity, úspěchy na SŠ a kvalita zaslaných esejí dovedly dle autora v předchozích výzkumech předpovídat nejen vysokoškolské GPA (nad rámec testů studijních předpokladů a známek ze SŠ), ale i další kritéria, jako úspěchy na VŠ a dosažené vedoucí pozice. Výstižně poznamenává, že osobnostní faktory umožňují zachytit, jak se uchazeč patrně v budoucnu zachová („will do“), z kognitivních schopností je možné vyčíst, jakých výkonů by uchazeč byl schopen („can do“) a dosavadní úspěchy ukazují, čeho již uchazeč dosáhl („have done“).

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?

Nehledě na výše uvedené informace je nejčastějším způsobem přijímání na vysokou školu u nás i v zahraničí ověřování tzv. „studijních předpokladů“. Stejně jako u akademického úspěchu se i v případě studijních předpokladů jedná o problematiku komplexní, která si zaslouží vlastní podkapitulu.

4.1 „Studijní předpoklady“ – co to je a jak je zjistit?

Jednotná definice „předpokladů ke studiu“ stejně jako „akademického úspěchu“ 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 (angl. *reasoning ability*), jejíž odlišnost od obecné inteligence je ovšem nejasná. Jiní (např. Kuncel et al. 2001) 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 škole³.

³ Tuto definici používají i některé velké univerzity v ČR, např. Masarykova univerzita v Brně (20. května 2013), Přírodovědecká fakulta UK v Praze (nedat.) či Univerzita Palackého v Olomouci (20. května 2013).

Počátky dnešních testů studijních předpokladů lze každopádně 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 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. 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: Testy předpokladů ke studiu používané na velkých univerzitách v ČR většinou zahrnují některé z následujících dovedností a znalostí: verbální myšlení, kritické myšlení, abstraktní uvažování, (logicko-) analytické myšlení, symbolické myšlení, numerické/kvantitativní myšlení, prostorová představivost a všeobecný/kulturní přehled (Masarykova univerzita v Brně, 20. května 2013; Přírodovědecká fakulta UK v Praze, nedat.; Univerzita Palackého v Olomouci, 20. května 2013)⁴. Společnost Scio (20. května 2013), jejíž test Obecných studijních předpokladů (OSP) je na VŠ také hojně využíván, od školního roku 2013/2014 do OSP zařadila také tzv. argumentační oddíl, jehož náplní by měla být především obsahová a logická analýza textů. Ještě před zavedením těchto změn při tom publikovala zjištění o souběžné validitě testů OSP a SAT na vzorku 104 uchazečů o studium na VŠ, a to jak v celkovém skóre ($r = 0,761$) tak mezi odpovídajícím oddílem verbálním (OSP) a

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

kritického čtení (SAT) ($r = 0,711$) a oddílem kvantitativním (OSP) a matematickým (SAT) ($r = 0,729$) (Scio, 15. ledna 2013).

Pod stejnými názvy subtestů různých testů studijních předpokladů se ovšem mnohdy skrývají jiné typy úloh; tak například verbální myšlení může sestávat z porovnávání synonym či antonym, ale i z otázek spíše lingvistických (např. slovtvorba). Podobně je tzv. všeobecný či kulturní přehled velmi širokou kategorií, jež může obsáhnout položky napříč různými obory – historie, zeměpis, společenské vědy aj. Tím se stává poměrně náročným na přípravu, jak z hlediska uchazečů, tak i z hlediska tvůrců testu, a můžeme se setkat s diskuzí nad jeho oprávněností (viz také kap. 4.2). 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 (srov. Matějů, Ježek, Műnich, Polechov, Slovk, Strakov, ... & Zrzav, 2009).

V tomto kontextu je moŹn zmnit studii Konenho, Basla a Mysliveka (2010), kteří hodnotili vliv pritomnosti oborovch testů a testů studijních predpokladů v prijmacm řzen na vysokou školu na podl a charakteristiky uchazeů, kteří jsou na danou školu prijati, a to na zklad dat Sondy Maturant 1998 a Uchaze 1998. Uvdj, Źe mnoho maturantů, kteří v testu studijních predpokladů (OSP) dosahuj stejnch či lepšch vsledků neŹ jejich spoluŹci, se nehls na vysokou školu, potenciln proto, Źe prijmac zkoušky byly postaveny na znalostnm/oborovm testu. Autoři tedy nasimulovali rozhodnutí o prijet na VŠ na zklad OSP a oborovch testů a došli k zveru, Źe pri oborovm testu maj vyšš šanci na prijet uchazei s vyššm socioekonomickm zzemm (operacionalizovanm jako vysokoškolsk vzdln otce, matky, a pritomnost potae v domcnosti) a z gymnzi, zatímco v modelu zaloŹenm na OSP jsou tato „zvhodnn“ niŹš. I pes diskutovan omezen dan simulace tedy autoři konstatuj, Źe prijmac řzen postaven na testech obecnch studijních predpokladů by mohlo bt „frovjš“ z hlediska soc. zzem uchazeů.

Avšak kromě toho, že neexistuje jednoznačná náplň různých testů předpokladů ke studiu, není také zcela jasné, nakolik se tyto liší od jiných kognitivně zaměřených testů (typicky od testů inteligence) – k tomuto tématu více v následující podkapitole.

4.1.1 Konstrukt „studijních předpokladů“

Jakkoliv někteří z tvůrců tvrdí, že se v případě testů studijních předpokladů nejedná o testy inteligence (např. Scio, 20. května 2013), typy úloh mohou mnohdy inteligenční testy připomínat, zvláště jedná-li se např. o doplňování číselných řad, mentální rotaci obrázků a podobně. Také Bollinger (2004) vznáší otázku, zda testy jako SAT měří jakousi vrozenou intelektovou kapacitu nebo spíše intelektové schopnosti, které se mohou časem zlepšovat a být naučeny. Upozorňuje, že ani sami tvůrci testů SAT, organizace College Board, nemají jednoznačnou odpověď: Ve svých popisech cílů testů SAT nejprve používají pojem „aptitude“ čili ověřování studijního potenciálu, později udávají coby jejich účel zjišťování schopnosti uvažovat (*reasoning ability*)⁵ a v další verzi již zkratka SAT zůstává bez významu a diskutuje se pouze „naučitelnost“ na testy (viz kap. 4.2).

Existuje tedy rozdíl mezi konstruktem „předpokladů ke studiu“ a inteligence? Gottfredsonová (2004) v této souvislosti hovoří o trendu upouštění od pojmu „inteligence“, na jehož místo nastupuje tzv. *obecná mentální schopnost* (g-faktor) učit se, usuzovat a řešit problémy. Zároveň upozorňuje na vysoké interkorelace mezi různými testy kognitivních schopností (verbální, matematické, prostorové), což podle ní nasvědčuje skutečnosti, že v pozadí všech z nich stojí jeden společný faktor. Tato interpretace je v souladu se zjištěními Carrolla (1993), který provedl metaanalýzu 450 studií zabývajících se různými kognitivními schopnostmi a na základě faktorové analýzy dospěl k třívrstvé, hierarchické teorii o kognitivních schopnostech: V nejvyšší vrstvě byl nalezen pouze jediný společný faktor, g. K němu se ve druhé vrstvě přidružují specializovanější schopnosti týkající se mj. řeči, prostorové představivosti, sluchového vnímání, paměti a rychlosti zpracování informací. Na nejnižší úrovni se pak nacházejí konkrétní složky či „sub-části“ těchto schopností, například porozumění textu nebo rozsah krátkodobé paměti. Všechny tyto

⁵ 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. Poněkud smutnější interpretace by byla, že eseje představují příliš časově zatěžující a na vyhodnocení náročný úkol, kterému se české vysoké školství raději vyhýbá.

složky spolu přitom dle Carrolla (1993) souvisí a jakýkoliv test určený k jejich ověřování zároveň z velké části měří právě společný faktor *g*.

Nabízí se tedy otázka, proč k přijímacímu řízení nepoužívat přímo standardizované IQ testy, zvláště když *g*-faktor je dle mnoha autorů nejlepším prediktorem akademického úspěchu (např. Gottfredson, 2004; Kuncel, Hezlett, & Ones, 2004 aj.), neboť pochopení vědeckých poznatků a jejich aplikace vyžaduje v první řadě dobré uvažování a analytické schopnosti. Zároveň existují i studie potvrzující vysokou korelaci mezi skóry testů studijních předpokladů a klasických IQ testů, jakým jsou například Ravenovy progresivní matrice. 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 ($r = 0.483$, $p < .001$) (po korekci omezené variability korelace stoupla na $r = 0.72$). Podobně společnost Scio (květen 2012) analyzovala 290 účastníků testu OSP, kteří zároveň vyplnili IQ test předkládaný společností Mensa ČR (blíže neidentifikovaný, leč formátem velmi podobný Ravenovým progresivním maticím) a zjistila korelaci $r = 0.5$.⁶ Ačkoliv tyto údaje každá ze skupin interpretuje poněkud jinak⁷, faktem zůstává, že schopnosti a vlastnosti měřené pomocí testů studijních předpokladů souvisejí určitým způsobem s obecnou inteligencí. Proč se tedy IQ testy neobjevují u přijímacích zkoušek?

Jedním z důvodů může být rozporuplné vnímání inteligenčních testů širokou veřejností. Rozsáhlou diskuzi a následný výzkum vyvolalo především nařčení IQ testů z neférovosti vůči některým společenským menšinám (viz kap. 6.2), ačkoliv v současnosti již mezi odborníky panuje shoda, že rozdílné výsledky různých skupin nejsou zaviněny předpojatostí testů jako spíše reálnými rozdíly ve výkonech těchto skupin (Gottfredson, 2004; Sackett, Borneman, & Connelly, 2008). Další námitkou může být nepřesnost bodového odhadu, s jakou se setkáme u většiny psychodiagnostických nástrojů. Statistici i odborníci sestavující příslušné manuály ostatně právě z tohoto důvodu doporučují využívání intervalů spolehlivosti a interpretaci naměřených skóre pouze s ohledem na příslušné pásmo výkonu, nikoliv na absolutní hodnotu. Je však těžké představit si seřazení

⁶ V článku o této studii však chybí některé důležité metodologické údaje nutné pro kvalitní interpretaci výsledků.

⁷ „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).

uchazečů pouze na základě intervalů spolehlivosti, do nichž svými skóry budou spadat, jelikož se nejspíše bude jednat o velmi podobné, ne-li shodné intervaly.

Na tomto místě je dobré poznamenat, že problematičnost bodového odhadu zasahuje i testy studijních předpokladů. Jak poznamenává Bollinger (2004), důraz, který rodiče a studenti přikládají každému bodu v testu SAT, neodpovídá jeho skutečným rozlišovacím schopnostem (rozdíl 10 bodů v tomto testu je z psychometrického hlediska poměrně nevýznamný); naopak spíše odráží velmi zkreslený systém hodnot v dnešní společnosti, který i tímto přístupem spoluvytváříme – totiž důraz na výsledky testu, nikoliv na skutečné schopnosti. Jakékoliv využití *hraničních hodnot* (angl. *cut-off scores*) v přijímacím procesu, pakliže se mu nemůžeme vyhnout, by tedy mělo být dobře odůvodněno a empiricky podloženo (Joint Committee for Standards et al., 1999).

V neposlední řadě může nepoužívání klasických inteligenčních testů v přijímacím řízení na VŠ opět souviset s jedním z pohledů na *férovost* tohoto procesu, totiž stejnou možností všech účastníků se na přijímací zkoušku připravit (viz kap. 4.2 a 6.2). Za problematické lze v tomto případě považovat dva body: Jednak vysoké riziko prozrazení podoby a tím devalvace výsledků inteligenčních testů, jednak vysoká genetická podmíněnost g-faktoru, která by tak potenciálně „znevýhodňovala“ některé uchazeče. Gottfredsonová (2004) poukazuje na konferenci Americké psychologické asociace v roce 1995, kde bylo konstatováno, že odhadovaná nejnižší hranice genetické komponenty inteligence je 40%. Současně ovšem upozorňuje, že rozvoj učení je do velké míry ovlivněn i prostředím, v němž se člověk nachází, a že schopnost učit se mají všichni lidé; proto je představa inteligence coby „elitářského“ nebo potenciálně diskriminujícího konceptu chybná. Vyšší IQ je pouze spojeno s lepší šancí učit se rychleji, samostatněji a i za pomoci abstraktních či komplexních instrukcí – což ostatně často studium na vysoké škole vyžaduje.

Zdá se tedy, že použití (pojmenování) testů studijních předpokladů je společensky přijatelnější než využití standardizovaných inteligenčních testů, ačkoliv je pravděpodobné, že ve výsledku měří ten samý faktor (viz Carroll, 1993). To ovšem nikterak neutišuje vlnu protestů i proti tomuto způsobu přijímání uchazečů ke studiu na vysoké škole, ačkoliv Bollinger (2004) poznamenává, že diskuze nad použitím testů typu SAT je spíše „hromosvodem“ strachu ze soutěživosti a hodnoty terciárního vzdělávání v dnešní společnosti.

4.2 Dosavadní výkon vs. výkonový potenciál a „naučitelnost“ na testy

V souvislosti s nejnovějšími revizemi SAT byl v zahraničí diskutován rozdíl mezi ověřováním „dosaženého úspěchu/výkonu“ (angl. *achievement*, měřící úroveň znalostí a opírající se o středoškolské kurikulum, tzn. co byl student schopen si odnést ze střední školy) a „schopností/potenciálu“ (angl. *aptitude*, snažící se zachytit méně obsahově závislé a více přenositelné kompetence, tzn. co vše by mohl zvládat v budoucnu) a jejich důležitostí 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). Bollinger (2004) poznamenává, že existuje zásadní rozdíl mezi tímto „připravováním se na testy“ a „učením se“ coby dlouhodobým procesem, kdy „dobrý student“ na střední škole automaticky ovládne všechny zjišťované dovednosti. Tento rozdíl dle autora ovšem, zdá se, často ignoruje nejen většina uchazečů, ale i středních škol, které se soustředí právě na přípravu svých studentů „na test“.

Zároveň Atkinson (2001) namítá, že „studijní potenciál“ není jasně definovaný ani měřitelný a jeho prediktivní validita (vzhledem ke známkovému průměru studentů v prvním roce) není nad rámec známek ze střední školy a oborových testů vysoká. Opírá se přitom o výzkum Geisera a Studleyho (2001) na vzorku téměř 78.000 studentů University of California, kteří zjišťovali relativní přínos (měřeno přírůstkem objasněné variance, R^2 , v regresním modelu) testů SAT I (obecné kognitivní schopnosti) nad rámec známek ze střední školy a oborových testů (SAT II). Oborové testy se v jejich analýzách dokonce ve všech třech měření ukázaly být nejlepším samostatným prediktorem, který vysvětloval 16% variance ve známkovém průměru studentů v prvním ročníku VŠ⁸. Na druhé straně najdeme i zastánce testů obecných studijních předpokladů argumentující právě faktem, že tyto testy jsou nezávislé na obsahu středoškolského kurikula, které se může napříč různými SŠ podstatně lišit (Matějů et al., 2009). Je však třeba podotknout, že takové testy by tedy de facto neměly obsahovat otázky z všeobecného přehledu (neboť tyto mohou s náplní a kvalitou výuky na SŠ silně souviset) a ani tak nebude zcela vyřešen problém s obsahovou definicí konstruktů studijních předpokladů.

⁸ Pro další výzkumy prediktivní validity testů studijních předpokladů a jiných prediktorů viz kap. 4.3.

Jedním z bodů v diskuzi o používání testů studijních předpokladů je také vliv „koučování“ čili nácviku testových otázek na konečný výsledek uchazeče, přičemž se spekuluje o znevýhodnění sociálně slabších skupin, které si nemohou dovolit placené cvičebnice a kurzy (NACAC, 2008). Sackett, Borneman a Connelly (2008) sice shrnují výsledky rozsáhlých metaanalytických studií se závěrem, že tato příprava konkrétní skóre uchazeče v „ostrém“ testu příliš neovlivní⁹, přesto se odvolávají na doporučení Zwickové (2007), totiž že nejsnazší obranou proti tomuto nařčení je bezplatné zpřístupnění cvičných testových materiálů všem potenciálním uchazečům. V souladu s tímto doporučením postupuje i česká legislativa, která ustanovuje povinnost zveřejňovat již proběhlá znění testů (viz kap. 6.2) (Vyhláška č. 343/2002 Sb.).

Pojďme se tedy podívat, jak jsou na tom testy studijních předpokladů a další prediktory vzhledem k předpovědím (různě operacionalizovaného) akademického úspěchu vysokoškolských studentů.

4.3 Prediktivní validita testů studijních předpokladů a dalších faktorů

Camara (2004a) mezi „tradiční“ prediktory akademického úspěchu řadí skóre testů SAT nebo ACT, známky ze střední školy, eseje a doporučující dopisy. Znamky ze střední školy přitom mohou pokrývat jednak specifické a všeobecné znalosti, verbální a matematické usuzování stejně jako jazykový projev, jednak i další faktory jako úsilí, svědomitost a docházku.¹⁰ Naproti tomu přijímací testy (SAT, ale i oborové) pokrývají pouze kognitivní dovednosti (vč. paměti), ač měřené za stejných podmínek. Camara (2004b) dále uvádí podrobné dělení prediktorů do pěti skupin, z nichž některé se obsahově překrývají navzájem:

- a) **Kognitivní měření.** V této kategorii se nacházejí již zmiňované známky ze střední školy, testy studijních předpokladů (SAT, ACT,...) nebo oborové testy.
- b) **Temperament.** Do této kategorie autor řadí osobnostní charakteristiky dle koncepce Big Five, sebepojetí, vůdčí schopnosti, zvědavost, komunikační dovednosti, motivaci a stanovené cíle, morální charakter, soucit, empatii a dobrovolnickou činnost.

⁹ Výjimkou u nedostatečně kvalitních testů by snad mohlo být získání tzv. „test-wiseness“ neboli schopnosti odhadnout správnou odpověď na základě jiných vodítek než skutečných znalostí, např. z délky jednotlivých možností v multiple-choice testu apod. Zkušební tvůrci testu by se však takovýchto nápověd měli být schopni vyvarovat (Schultz, & Whitney, 2005).

¹⁰ Nesmíme ovšem zapomínat na možná zkreslení odlišnou „známkovací praxí“.

- c) **Osobnostní kvality a zkušenosti.** Operacionalizací této skupiny prediktorů mohou být různé doporučující dopisy, eseje, životopis, pohovory, akademická ocenění, mimoškolní aktivity a výjimečné výkony v oblasti umění, sportu nebo vědy, ale i plynulá znalost cizího jazyka.
- d) **Charakteristiky uchazeče.** Tato kategorie v podstatě odpovídá demografickým ukazatelům, jako je pohlaví, socioekonomický status, příslušnost k etnické menšině, věk, současné zaměstnání apod.
- e) **Jiné.** Sem autor řadí faktory spíše situační, tzn. kvalita střední školy, ze které se uchazeč hlásí, složení aktuálního ročníku všech uchazečů a podmínky na dané VŠ, doporučení přicházející „shora“ aj.

Podobné rozdělení najdeme i u Richardsona, Bonda a Abrahama (2012), kteří provedli metaanalýzu studií na velkých vzorcích vysokoškoláků ($N > 1.000$) zabývajících se predikcí vysokoškolského GPA od roku 1997 a rozdělili sledované prediktory do sedmi kategorií:

- a) **demografické faktory** (pohlaví, věk, socioekonomický status);
- b) **tradiční testy kognitivních schopností či dosažených výsledků** (inteligenční testy, SAT/ACT, známky ze SŠ);
- c) **osobnostní vlastnosti** (dle Big Five a dále prokrastinace, potřeba intelektuální činnosti a emoční inteligence);
- d) **motivační faktory** (jako místo kontroly, optimismus, akademická a výkonová vnímaná vlastní zdatnost a orientace na cíl);
- e) **seberegulační strategie** (testová anxieta, výdrž, time management apod.);
- f) **přístup k učení** (povrchní, hloubkový či strategický); a
- g) **kontextuální vlivy** (jako stres, podpora okolí a integrace do akademického prostředí).

Jako nejsilnější samostatné prediktory známek na VŠ se ukázaly známky ze střední školy ($r = 0,4$), testy SAT/ACT ($r = 0,29-0,4$), vnímaná vlastní zdatnost ($r = 0,31-0,59$), orientace na cíl ($r = 0,35$) a výdrž ($r = 0,32$). Kombinace těchto prediktorů v regresní analýze poté vedla k objasnění 28% variance kritéria. Willingham (18. února 2013) přitom komentuje, že statistická významnost těchto výsledků je z velké části dána velikostí zkoumaných

vzorků a také v případě sebeposuzovacích měřítek (vnímaná vlastní zdatnost, orientace na cíl, výdrž) možným překryvem, resp. zkrácením jednotlivých konceptů.

Ve výzkumné praxi je coby kritérium predikční validity nejčastěji používán právě výše zmíněný známkový průměr (jak v prvním roce studia - FYGPA, tak i za studium celkově - GPA) a nejvíce studií nalezneme k prediktivní schopnosti testů studijních předpokladů a středoškolských známek. Camara (2004a, 2004b) dokonce poznamenává, že od roku 1996 do r. 2002 proběhlo 987 validizačních studií, organizovaných výzkumným oddělením College Board (College Board validity service), v nichž byl test SAT použit coby prediktor. Pouze 12 z těchto studií však využilo jako predikovaná kritéria kromě FYGPA také známky z jednotlivých předmětů; nalezené korelace mezi SAT a FYGPA se při tom pohybovaly v rozmezí přibližně $r = 0,44-0,62$. Sackett, Borneman a Connelly (2008) přitom upozorňují, že pokud by jimi pozorovaná korelace mezi výsledkem SAT a FYGPA u 165,000 studentů na 41 vysokých školách ($r = 0,47$ po korekci pro omezení variability) byla založena pouze na datech z konkrétních předmětů, jimiž prošli všichni studenti (data dostupná od 147,000 studentů), dosahovala by odhadem hodnoty $r = 0,55$ (viz kap. 3).

Pokud jde o vzájemnou souvislost testů studijních předpokladů a středoškolských známek, 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 $r = 0,44$ ¹¹. K podobným výsledkům došli i Kobrin, Patterson, Shaw, Mattern a Barbuti (2008), kteří ověřovali validitu testu SAT po zavedení některých změn ve složení a obsahu jednotlivých subtestů (kritické čtení, matematika, psaní) v roce 2005. Na vzorku více než 150 tisíc studentů přijatých celkem na 110 čtyřletých vysokých škol se jako nejsilnější samostatný prediktor FYGPA ukázaly být známky ze střední školy (po korekci pro omezení variability $r = 0,54$). Ze subtestů SAT poté nejvyšší korelace s FYGPA dosáhla nově přidaná část psaní eseje (korigované $r =$

¹¹ 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 $r = 0.41-0.64$.

0,51), i když tato hodnota se příliš nelišila od predikce na základě matematického subtestu (korigované $r = 0,47$), kritického čtení (korigované $r = 0,48$) a celkového skóru ze všech subtestů (korigované $r = 0,53$). Nejlepší predikci FYGPA celkově zajistila kombinace známek ze SŠ a všech subtestů SAT (korigované $r = 0,62$), ačkoliv přínos subtestu psaní nad rámec kombinace známek ze SŠ, SAT-matematického subtestu a SAT-kritického čtení (korigované $r = 0,61$) lze považovat za minimální. Nabízí se tedy otázka, zda je možné přínos testu SAT 0,08 nad rámec známek ze střední školy považovat za dostatečný argument pro použití testů studijních předpokladů v přijímacím řízení, nebo zda by naopak nebylo snazší opírat se pouze o známky ze SŠ (se všemi jejich potenciálními nevýhodami), doplněné např. jinými (ne)kognitivními kritérii. Také Richardson, Bond a Abraham (2012) uvádějí, že středoškolské známky jsou lepším prediktorem VŠ známek než testy SAT/ACT, ačkoliv vysvětlují rozdílný podíl variance (dohromady přibližně 22-25%)¹².

Co se týče jiných prediktorů než středoškolských známek a testů studijních předpokladů, Sternberg (2004) navrhuje rozšíření současného způsobu ověřování předpokladů ke studiu o další dimenze. Opírá se při tom o svou teorii tzv. úspěšné inteligence, kterou definuje jako schopnost dosáhnout úspěchu v životě vzhledem k vlastním standardům a sociokulturnímu kontextu. Tento koncept sestává z vyvážených *analytických schopností*, *Kreativit* a *praktických schopností*, vztahujících se k řešení běžných denních problémů. Za účelem jeho zjišťování u studentů vyvinul autor baterii zvanou Rainbow Measures, kterou předložil 777 studentům a jejíž pomocí se pokoušel předikovat jejich celkový GPA. Z jeho analýz prostřednictvím hierarchické regrese vyplývá několik poznatků:

Zprv, při současném zadání známek ze střední školy a výsledků SAT jsou známky ze SŠ výrazně lepším prediktorem než SAT a dohromady vysvětlují 16,4% celkové variance. Přidání výsledků z analytické části Rainbow Measures predikci nezlepší, spíše se jednotlivé podíly „rozloží“ mezi všechny složky, přičemž nejlepším prediktorem zůstávají známky ze SŠ. Zadruhé, přidáním výsledků z praktické části se podíl vysvětlené variance zvedne na 16,6% a signifikantními prediktory jsou právě praktická inteligence a známky ze SŠ. A konečně po zadání oddílu kreativity stoupne vysvětlená variance na 24,4%, přičemž signifikantním prediktorem zůstávají známky ze SŠ a skóry kreativity (skóry praktické inteligence ztrácí na významu).

¹² K českým výzkumům na toto téma více níže.

Vzhledem k faktu, že podíly objasněné variance vysokoškolského GPA pomocí tohoto modelu se příliš neliší od nálezů dalších zahraničních autorů (viz níže) stran prediktivní validity testu SAT a středoškolských známek, stojí za úvahu, zda by zavedení Sternbergova testu bylo pro současnou praxi přínosem. Jak jeho autor uvádí, vyplnění jednotlivých částí by mělo zabrat maximálně 70 minut (faktická doba vyplňování se ve výše zmíněné studii pohybovala mezi 2 a 4 hodinami). Tento časový údaj se výrazně neliší od doby potřebné pro vyplnění testů SAT, na české poměry se ovšem může jednat o dobu poměrně dlouhou. Za zvážení však stojí přínos kreativní části této baterie k celkové predikci; v budoucím výzkumu by mohlo být zajímavé zaměřit se na tuto část blíže.

Z českých výzkumů na téma prediktivní validity různých proměnných vzhledem k úspěchu na vysoké škole 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 mj. 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, 339).

Obdobně 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 regresí vyplynulo, že nejlepšími prediktory celkového známkového průměru za 3 roky studia jsou výsledky oborového testu z fyziky a průměr známek z fyziky ze střední školy, které společně vysvětlovaly 21% variance, doplněné o hodnocení studentovy motivace a

výsledek škály attachmentu (dohromady vysvětlující 32% variance). 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.

Podobným výzkumům se věnovali i Štuka a Šimeček (2006), kteří za pomoci logistické regrese analyzovali prediktory úspěšného ukončení studia medicíny na 1.LF UK v Praze u 626 studentů zapsaných v letech 1994-1997. Mezi sledované proměnné patřilo „pohlaví, typ středoškolského vzdělání (gymnaziální/jiné), počet bodů na jednotlivých částech přijímacích zkoušek (biologie, chemie, fyzika), průměrné známky za středoškolské studium z matematiky, fyziky, chemie, biologie a českého jazyka a průměrné známky z jednotlivých ročníků. Z těchto údajů byl dále dopočten součet bodů za přijímací zkoušky celkově, průměr z průměrů za jednotlivé ročníky a rozdíl průměrů v posledním a prvním ročníku střední školy zrcadlící zlepšení či zhoršení žáka během studia“ (Štuka, & Šimeček, 2006, 2). Na základě shlukové analýzy autoři konstatují, že známky ze SŠ nesou poněkud odlišnou informaci než známky z přijímacích zkoušek, a zdají se být také lepšími samostatnými prediktory úspěšného dokončení studia. Jako nejlepší a zároveň nejuspěšnější kombinace prediktorů se ukázal model „[s výsledky] přijímacího testu z fyziky a chemie, průměrné středoškolské známky z biologie a chemie a rozdílu mezi průměrem v čtvrtém a prvním ročníku střední školy“ (Štuka, & Šimeček, 2006, 4). Autoři tedy doporučují zohlednit při přijímacím řízení i známky ze SŠ.

Štuka (2012) se k této problematice vrací ve své disertační práci, kdy 1.LF UK v roce 2002 umožnila přijetí části studentů na základě jejich prospěchu na střední škole a zároveň v roce 2008 zavedla vedle testů z chemie, biologie a fyziky (a následného ústního pohovoru) ještě test všeobecné studijní připravenosti (VSP). Otázky pro „odborné“ testy si 1.LF tvoří sama a dle vyjádření autora se při tom potýká s podobnými problémy jako FF UP, totiž s jejich nízkou „životností“, personálně náročnou a finančně nepříliš zvýhodněnou tvorbou, variabilní validitou i reliabilitou jednotlivých otázek (jež je nemožné pilotně testovat) atp. Test VSP obsahuje 60 otázek zaměřených na porozumění odbornému textu a grafům, prostorovou představivost, logické uvažování a schopnost

postřehnout detaily; byl vytvořen na FF UK a dle autora „byl zaměřen na intelektuální schopnosti, a to zejména na logické a analytické schopnosti a prostorovou představivost. Testové otázky formou i obsahem odpovídají otázkám z IQ testů“ (Štuka, 2012, 18)¹³.

Kritériem „úspěšnosti“ studia bylo při posuzování vlivu středoškolských známek (383 studentů přijatých v roce 1999) dokončení studia. Prediktory přitom byly informace o typu SŠ a maturitním roce, průměry známek z matematiky, fyziky, chemie, biologie a českého za 4 roky i celkový průměr známek za studium, a dále body z přijímacích testů z biologie, chemie a fyziky a jejich součet. Stejně jako v předchozím článku ukázala shluková analýza více podobností mezi známkami ze SŠ navzájem a známkami z odborných testů zvláště, než mezi těmito prediktory např. oborově (SŠ-biologie a test-biologie)¹⁴. Skóre odborných testů v modelu logistické regrese vysvětlovala 13% variance kritéria, celkový průměr známek na SŠ samostatně vysvětloval 15% variance a kombinace obou prediktorů (s váhami 0,45 pro odborné testy a 0,55 pro známky ze SŠ) objasňovala 22% variance. Jako nejvhodnější k predikci úspěšného dokončení studia se poté jevil model s typem školy a rokem maturity, středoškolským průměrem a testovými skóry v poměru 0,09 : 0,1 : 0,35 : 0,46, který vysvětloval 31% variability kritéria. Autor tedy shrnuje, že zahrnutí známek ze střední školy do procesu přijímacího řízení se zdá více než vhodné. Současně ale upozorňuje, že uchazeči přijatí pouze na základě známek měli tím vyšší tendenci k zápisu na jiné fakultě než 1.LF UK (kam byli také přijati), čím lepší jejich prospěch byl; tento trend se u uchazečů s absolvovanou přijímací zkouškou neprojevil. Je tedy možné se domýšlet, že i přes zohlednění středoškolského prospěchu je konání přijímací zkoušky dobré z hlediska zvýšení „hodnoty“ přijímacího procesu v očích uchazečů (a tudíž menší míry „úspěšných odpadlíků“), přičemž odborné testy v tomto řízení hrají podstatnou roli.

Pokud jde o význam testů VSP, zde byl coby kritérium zvolen průměr známek v prvním roce studia (FYGPA) a fakt, zda uchazeč studia zanechal do 1 roku po zápisu či nikoliv. Korelace celkového skóru z odborných testů s FYGPA byla vyšší ($r = 0,3$; $p < 0,001$) než korelace testu VSP ($r = 0,13$; $p < 0,001$) a při dichotomickém kritériu zanechání studia zůstala významnou pouze korelace s odbornými testy. Štuka (2012) udává, že tento „sporný přínos“ VSP je v souladu s poznatky zahraničních výzkumů, které využití testů studijních předpokladů na lékařských fakultách nad rámec oborových testů či

¹³ Srov. diskuzi o konstruktu studijních předpokladů v kap. 4.1.1.

¹⁴ To by mohlo vyvolat zajímavou diskuzi o původu těchto rozdílů.

maturitního průměru příliš nedoporučují. Podobně Táborská, Mužík, Jarkovský, & Weissová (2005, in Žoudlík, 2009) uvádějí u uchazečů o studium na LF MU v Brně v roce 2003 pozitivní korelace mezi výsledky přijímacích zkoušek a známkovým průměrem vybraných (klíčových) předmětů v prvním roce studia, přičemž oborové testy (fyzika, biologie a chemie) se ukazují z hlediska souvislosti s GPA významnější než výsledky TSP.

Na závěr uvedme ještě informace k Testu studijních předpokladů (TSP) používanému na Masarykově univerzitě v Brně, jak je předkládá Žoudlík (2009). Ten se ve své magisterské práci věnoval srovnání různých typů přijímacích zkoušek a jejich prediktivní validity vzhledem ke známkovému průměru na VŠ či jejímu úspěšnému dokončení a aplikaci těchto poznatků v prostředí MU. Jednalo se čistě o práci teoretickou, přesto z ní lze získat přehled (shrnutý v *Tab. 1*) základních charakteristik TSP používaného na MU. Tyto hodnoty mohou sloužit jako orientační vodítka k posouzení testů SPF na FF UP.

Tab. 1: Základní charakteristiky TSP MU v Brně (dle Žoudlík, 2009)

	2003	2004	2005	2006	2007	2008	2009	2010
Počet variant	24	24	24	18	18	18	18	18
Počet úloh	80	80	80	80	64	70	70	70
Počet subtestů	8	8	8	8	8	7	7	7
Časový limit	80	80	80	80	80	90	90	100
Prům. počet správných odp.	X	47,18	46,72	44,71	26,7	26,9	26,1	X
Směr.odchylka	X	9,62	8,95	8,4	7,2	7,1	7,8	X
Mín. spr. odp.	X	X	X	4	1	0	0	X
Max. spr. odp.	77	78	80	73	57	65	61	X
Cronbach. alfa	X	X	X	0,784	0,727	0,757	0,776	X

4.4 Doporučení pro vysoké školy k ověřování předpokladů ke studiu

S ohledem na výše uvedené poznatky je možné formulovat několik doporučení pro vysoké školy, jejichž aplikace v praxi by mohla přispět ke zkvalitňování procesu přijímacího řízení. V první řadě se lze řídit metodikou vytvořenou Lairdem (2004) na základě pokynů amerického ministerstva školství k užívání standardizovaných testů

v situacích, kdy následné rozhodnutí může výrazně ovlivnit budoucí život prověřovaného člověka (tzv. *high-stakes decision making*):

- 1) **Univerzita by měla jasně definovat účel svého přijímacího řízení a stanovených požadavků na uchazeče**, přičemž tyto by měly přímo souviset s deklarovaným posláním univerzity. Pokud je tedy např. cílem univerzity nabrat třídu s nejvyšším možným známkovým průměrem v prvním roce, může odůvodnit použití testů studijních předpokladů coby velmi dobrého prediktoru FYGPA. Cíle přitom mohou být různé, od výběru takových uchazečů, kteří nejlépe využijí zdrojů fakulty, přes třídu, která bude nejlépe odrážet populační rozložení ve spádové oblasti až po výběr uchazečů s největším potenciálem pomáhat ve své budoucí kariéře ostatním, atp.
- 2) **Univerzita by vzhledem ke svým deklarovaným cílům měla v přijímacím řízení využívat vhodné prediktory**. Tím je myšlena např. situace, kdy je cílem vysoké školy nabrat ročník potenciálních filantropů, ovšem místo ověřování dosavadních mimoškolních dobrovolnických aktivit uchazečů by byl výběrovým kritériem např. průměr známek ze střední školy. Nejde tedy o prediktory samotné – mohou jimi být známky či předměty absolvované na SŠ, testová skóre, dosažené úspěchy, talent, motivace či schopnost překonávat překážky – ale o jejich vhodnost s ohledem na zamýšlené cíle přijímacího řízení. Laird (2004) také upozorňuje, že veškeré faktory by neměly být brány jako „statické“ (ve smyslu „uchazeč splňuje/nesplňuje, a tudíž ne/dostane body“) či jako pouhé součásti indexu k ne/přijetí uchazeče, nýbrž být pokud možno posuzovány individuálně s ohledem na jejich kontext.
- 3) **Pokud univerzita připisuje jednotlivým prediktorům různé váhy, mělo by být zřejmé, na čem se tyto váhy zakládají**. Určení vah může probíhat buď na základě (propracované) teorie, nebo empiricky, např. za použití vícenásobné regrese. Zde je ovšem opět otázkou, co přesně má být predikováno. Zároveň je třeba dát si pozor a pravidelně ověřovat, zda váhy nezastarávají: Je kupříkladu snadné si představit, že váha 0,5 u známkového průměru ze střední školy bude mít smysl, když bude populace uchazečů v tomto ohledu dostatečně rozmanitá, začne-li ovšem většina žadatelů o místo na dané VŠ dosahovat průměru 1,0, bude lépe tuto váhu snížit.

- 4) **Univerzita by své prediktory měla pravidelně evaluovat z hlediska toho, zda pomáhají dosahovat deklarované cíle.** Jedná se tedy o nabádání k dobré praxi pravidelných validizačních studií; praktická úskalí v našich podmínkách diskutujeme níže.
- 5) **Univerzita by ve svém přijímacím procesu pokud možno neměla používat vzorce nebo hrubá dělicí (*cut-off*) skóre.** Laird (2004) upozorňuje, že veškeré matematické či statistické vzorce svádějí k falešnému přesvědčení o vlastní bezchybnosti (podobně jako formulace „vědecké výzkumy dokazují“) a mohou tak zanedbávat řadu zkreslení (viz např. bod 3). Vzorce, a především dělicí skóre však mohu mít i závažnější dopady, a to pokud jediný bod bude moci rozhodovat o přijetí či nepřijetí uchazeče. Pak je totiž na místě zjišťovat, zda test, v němž tento bod chyběl, byl skutečně zkonstruován s tak precizní rozlišovací schopností položek, že jeden bod i v praxi znamená významný rozdíl v dovednostech uchazeče. Jakkoliv tedy použití vzorců usnadňuje a urychluje výběr z velkého množství uchazečů za minimálních personálních a finančních nákladů, z hlediska kvality přijímacího řízení by měly být vždy posuzovány v kontextu možností a zázemí každého jednotlivce.
- 6) **Univerzita by si měla pročíst co nejvíce individuálních přihlášek a portfólií.** V případě FF UP by se tedy měl vyžadovat např. životopis, motivační či doporučující dopisy, potvrzení o praxích, dosažených úspěších či publikovaných pracích v oboru a mimo něj apod., jako je tomu v případě zápisu prostřednictvím institutu celoživotního vzdělávání. I Laird (2004) připouští, že takový postup je z hlediska času i potřebného personálu velmi náročný a může být napadnutelný coby vysoce subjektivní, vidí však jeho přínos ke kvalitě výběru z uchazečů. Hodnotitele je navíc možné předem vyškolit a ověřit si jednotnost jejich měřítek (v psychometrice známé také jako *inter-rater reliability*) a jejich konečná rozhodnutí ještě nezávisle přezkoumat, což by mohlo výsledku ubrat na subjektivitě.
- 7) **Univerzita by při hodnocení úspěchů uchazečů měla zohledňovat i jejich sociokulturní zázemí.** Toto doporučení se prolíná s předchozími body; na tomto místě tedy pouze uvedeme, že se nemá jednat o „zvýhodnění“ uchazečů z nějaké specifického sociálního či kulturního prostředí, jako spíše o zohlednění možností v lokalitě, v níž uchazeč vyrůstal. Například studentovi střední odborné školy v malém městě nemusely být dostupné stejné dobrovolnické aktivity jako studentovi z velkoměsta, a jejich ne/vykonávání by tedy mělo být posuzováno

jinak. Na druhou stranu i zde můžeme argumentovat různorodou definicí „férovosti“.

- 8) **Univerzita by se měla snažit dozvědět co nejvíce o středních školách, z nichž se její uchazeči hlásí.** Jakkoliv byl tento bod formulován spíše v souvislosti se vzdělávací praxí v USA, může být relevantní i v českém prostředí – například v souvislosti s předchozím bodem či s rozdílnou praxí známkování na různých školách, popř. u různých učitelů.
- 9) **Univerzita by si měla vést záznamy o výkonech uchazečů z jednotlivých SŠ, kteří byli zapsáni v předchozích letech, a zohlednit tuto informaci při přijímání uchazečů ze stejné SŠ.** I toto doporučení je intuitivně srozumitelnější v kontextu USA, kdy se z jedné střední školy na danou VŠ hlásí i stovky, nejen desítky uchazečů, může však dopomoci k lepší prediktivní validitě – kterou je, samozřejmě, i tak potřeba nejprve ověřit.
- 10) **Univerzita by si měla, jak jen to je možné, ověřovat informace, které jí studenti předkládají.** Na problém verifikovatelnosti údajů narazila i FF UP, která v minulosti vyžadovala již v elektronické přihlášce např. známky z posledního pololetního vysvědčení na střední škole a další data o dosavadním studiu. Bez doložení kopií těchto vysvědčení však nebyl způsob, jak si (efektivně a úsporně) tyto informace ověřit, a proto se od povinnosti jejich uvádění upustilo – jakkoliv by se nám nyní tato data hodila k lepší validizaci přijímacího procesu. Laird (2004) ovšem u tohoto bodu zmiňuje i skutečnost, že by pro školy bylo dobré vědět, zda a případně kde a jak se uchazeči na přijímací řízení připravovali (např. jaké podstoupili kurzy či cvičné testy), což by mohlo usnadnit zhodnocení dopadu „učení se na test/přípravy na pohovor“ na jeho faktické výsledky. Z našeho pohledu se jedná o velmi zajímavou myšlenku, kterou by bylo vhodné v budoucnosti blíže prozkoumat.

Zadruhé, co se týče ověřování validity přijímacího procesu (bod 4), 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 podobě¹⁵, jako např. známky z maturitního vysvědčení, dokládané mnohdy až u zápisu ověřenou kopií dokumentu¹⁶) a jejich spárování při zachování anonymity jedince a bezpečnosti dat může být poměrně značným logistickým oříškem. Jedním z doporučení pro FF UP by tedy mohlo být vytvoření „multidimenzionální“ databáze informací o studentech: Zde by mohly být uloženy údaje z jejich el. přihlášek a maturitních vysvědčení, záznamy o průběhu studia (absolvované předměty a prospěch) případně i o jejich dalších aktivitách – zapojení se do studentských či dobrovolnických organizací, akademického senátu nebo zájmových kroužků; účast na konferencích či kurzech nabízených nad rámec běžné výuky atp. Tyto informace by mohly být evidovány buď prostřednictvím studijního oddělení, případně by je do určité míry mohli studenti vyplňovat sami (např. co se týče mimoškolních aktivit). Podobná databáze by poté mohla sloužit k pravidelným validizačním studiím přijímacího řízení, jako je tomu dobrou praxí v zahraničí.

Zatřetí by mohlo být i pro FF UP žádoucí zamyslet se nad alternativními způsoby ověřování studijních předpokladů, resp. dalších požadavků na uchazeče o studium na FF UP. Minimálně pro zhodnocení výkonových komponent ve Sternbergově pojetí již současná psychologie disponuje standardizovanými nástroji. Pomineme-li časově náročnou testovou baterii vytvořenou samotným Sternbergem (2004), analytické schopnosti kupříkladu většinou ověřují inteligenční testy¹⁷. Existují i testy kreativity a inventáře situačního úsudku (praktické inteligence). Například Schmitt, Oswald a Gillespie (2004) vyvinuli 12-dimenzionální škálu s multiple-choice otázkami pro vysokoškolské studenty, v níž mohou posuzovat své předpokládané reakce v situacích spojených s vysokoškolským životem (např. zda by šli na výstavu v rámci kurzu, na který běžně nechodí, apod.). Pokud jde o dělení dle Campbella et al. (1993, in Camara, 2004a), deklarativní vědomosti by se daly zjišťovat oborovými testy, přičemž podobně by mohl být navržen test k ověření určitých procedurálních znalostí; je však otázkou, co by takový test měl zahrnovat. Pro zhodnocení motivace uchazečů by pak byl patrně nejlepší ústní pohovor, ačkoliv jsou k dispozici i motivační dotazníky – zde však předpokládáme vyšší riziko zkreslení

¹⁵ Také Camara (2004a) uvádí, že i v zahraničí jsou často informace o potenciálních prediktorech akademického úspěchu uchazečů k nalezení pouze v papírové formě, zatímco informace o známkách či absolvovaných předmětech se nacházejí v samostatných elektronických databázích, což může znesnadňovat jejich propojování.

¹⁶ 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.

¹⁷ Jejich využití při přijímacích zkouškách diskutujeme v kap 4.1.1.

vzhledem k tomu, že je většinou celkem jasné, na co se dotazník ptá, a lze předpokládat tendenci uchazeče prezentovat se v co nejlepším světle (např. Camara, 2004b). Z tohoto výčtu by však mělo být zřejmé, že již existují nástroje pro zhodnocení více než jedné složky výkonu, která by mohla predikovat „úspěšné studium“ na vysoké škole.

Problémem zavedení těchto opatření v praxi může být, že v mnoha případech (snad až na talentové zkoušky) se testování (rozumějme zjišťování pomocí psychodiagnostických metod) jakýchkoliv jiných vlastností osobnosti, než jsou kognitivní schopnosti, jeví jako obtížně prosaditelné. Důvodem pro tento stav mohou být obavy z napadení regulérnosti přijímacího řízení či osočení z diskriminace. Také Camara (2004b) poznamenává, že akceptace osobnostních dotazníků veřejností (resp. v případě uchazečů o zaměstnání) je nižší než akceptace kognitivních testů, přičemž nejlépe (co by relevantní) jsou vnímány osobní pohovory. To je sice dobrým znamením pro obory, které této metody využívají k užšímu výběru kandidátů v současnosti, nicméně to nemění nic na faktu, že členové přijímacích komisí daných oborů se tak mohou spoléhat jen na svůj kvalifikovaný odhad a úsudek. Tím může v některých případech docházet k subjektivnímu zkreslení (viz také Žoudlík, 2009). Je třeba mít také na paměti, že veškeré zvolené prediktory by měly splňovat standardní kritéria kvality (viz kap. 6) a další požadavky kladené na nástroje používané v tzv. *high-stakes* situacích (srov. Joint Committee for Standards et al., 1999).

5 Test předpokladů ke studiu na FF UP (SPF)

Zatím poslední významná změna podoby přijímacího řízení na Filozofické fakultě UP proběhla v roce 2011. V té době se ustoupilo od využívání testů společnosti Scio a kromě oborových testů a ústních přijímacích zkoušek FF UP zavedla vlastní Test předpokladů ke studiu (SPF). Ten je od ak. roku 2011/2012 povinný pro všechny uchazeče o bakalářské studium, a to jak v prezenční, tak v kombinované formě. Na rozdíl od oborového testu, jehož obsah se orientuje na oborově specifické znalosti a dovednosti (podobně jako ústní pohovor, který mj. ověřuje ještě motivaci ke studiu), má SPF zjišťovat obecnou schopnost „úspěšně studovat na FF UP v Olomouci“ (Grigárková, 18. října 2013, 1).

Tato běžná, byť vágní definice (viz. kap. 4.1) je dále operacionalizována výčtem charakteristik, které má test ověřovat, totiž verbální myšlení, kritické myšlení, logicko-analytické myšlení (zde lze od r. 2012/2013 identifikovat oddíl prostorové představivosti) a všeobecný, resp. společensko-humanitní přehled. Celkově test obsahuje 50 otázek, na jejichž vyplnění má uchazeč 60 minut, přičemž z pěti odpovědí je vždy pouze jedna správná. Ačkoliv jsou otázky různě časově náročné, jejich bodové hodnocení je stejné, přičemž za špatnou či žádnou odpověď se body neodečítají (Opletalová, & Pavelková, 2013). V letech 2011 a 2012 bylo vždy administrováno celkem 16 verzí testu, v roce 2013 již jen 12 verzí, přičemž pokaždé byly dvě verze paralelní (obsahovaly stejné otázky, pouze v jiném pořadí a/nebo odlišné označení správných odpovědí)¹⁸. Základní přehled o složení testů v uplynulých třech ročnících podává *Tabulka 2*; podrobnější popis bude k nalezení v plánované publikaci „Tvorba a administrace testů studijních předpokladů“ (Charvát, Opletalová, Tošenovská, Viktorová, & Vobořil, 2014).

V případě, že se uchazeč hlásí na více oborů, absolvuje SPF pouze jednou a výsledek se mu započítá ke všem přihláškám; pokud se uchazeč hlásí na dvouobor, musí získat dostatečný počet bodů na každém oboru, jinak není ke studiu přijat. Vyhodnocení testů SPF probíhá prostřednictvím externí firmy, která anonymně (pouze pod přiděleným univerzitním číslem) zpracovává odpověďové archy a zpětně poskytuje fakultě informace

¹⁸ Viz kapitola o reliabilitě paralelních forem.

Tab. 2: Skladba testů SPF v letech 2011-2013

Oddíly testu	SPF 2011	SPF 2012	SPF 2013
Verbální myšlení (celkem)	7	15	15
z toho:			
<i>Antonyma</i>	1	1	1
<i>Archaismy</i>	X	1	1
<i>Frazeologie obecně</i>	1	X	X
<i>Frazeologie (ustálené výrazy)</i>	X	1	1
<i>Frazeologie (větné frazeologismy)</i>	X	1	1
<i>Hláskosloví/grafika, mezijazyková homonymie, porozumění slov, syntax, stylistika</i>	X	5	5
<i>Slovotvorba obecně</i>	1	X	X
<i>Slovotvorba – české komponenty</i>	X	1	1
<i>Slovotvorba – cizí komponenty</i>	X	1	1
<i>Synonyma českého původu</i>	1	1	1
<i>Synonyma (cizí slovo)</i>	1	1	1
<i>Syntagmatické a paradigmatické vztahy</i>	1	1	1
<i>Význam slov, příslušnost k sémantickému poli</i>	1	1	1
Kritické myšlení (celkem)	6	3	3
Logicko-analytické myšlení (celkem)	7	12	12
z toho:			
<i>Numerické výroky</i>	?	6-7	6-7
<i>Prostorová představivost</i>	X	4	4
<i>Výroky</i>	?	1-2	1-2
Všeobecný přehled (celkem)	30	20	20
z toho:			
<i>Dějiny umění</i>	2	2	2
<i>Ekonomie</i>	3	3	3
<i>Filologie</i>	5	3	3
<i>Filozofie</i>	2	2-3	2-3
<i>Historie a geografie</i>	6	2-3	2-3
<i>Literárněvědná terminologie, teologie</i>	0-1	0-1	0-1
<i>Literatura česká (vč. divadla)</i>	3	2	2
<i>Literatura světová (vč. divadla)</i>	3	2	2
<i>Politologie</i>	3	X	X
<i>Sociologie a psychologie</i>	3	3	3

Pozn.: ?= počty nejsou uvedeny; X = tento typ otázek nebyl použit. Tabulka je založena na analýzách Opletalové & Pavelkové (2013).

o dosaženém počtu bodů a percentilovém hodnocení uchazeče¹⁹ (Grigárková, 18. října 2013, a dále osobní sdělení). Tato firma ovšem neprovádí položkovou analýzu ani ověření dalších kritérií kvality testu, což byl jeden z důvodů iniciace projektu IGA FF_2013_011.

Před samotným popisem analýz je třeba ještě uvést, že na některé obory je kromě testu SPF předepsán navíc oborový test (OT) a/nebo ústní pohovor (UZ), jež lze v některých případech prominout na základě doložených znalostí (jazykový certifikát, maturitní vysvědčení apod.). Oborový test a/nebo ústní zkouška se konají v tentýž den jako testy SPF (neslouží tudíž coby „prvotní síto“), s výjimkou oboru psychologie v roce 2011, oborů psychologie a angličtina se zaměřením na tlumočení a překlad v roce 2012 a oborů andragogika v profilaci na personální management, psychologie a angličtina se zaměřením na komunitní tlumočení a překlad v roce 2013 (Grigárková, 18. října 2013; Zpráva o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2011/2012, nedat.; Zpráva o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2012/2013, nedat.). V těchto případech postoupili k ústnímu pohovoru pouze nejméně úspěšní uchazeči. Pakliže je konečný verdikt o (ne)přijetí uchazeče závislý na kombinaci testu SPF a výsledku oborového testu/ústního pohovoru, je celkový skóre počítán jako 30% dosažených bodů v SPF + 70% dosažených bodů v OT/UZ²⁰; tento poměr plánujeme zohlednit i v našich analýzách. Závisí-li přijetí uchazeče pouze na výsledku testu SPF, liší se ještě mezioborově bodové hranice pro (ne)přijetí, přičemž některé jsou benevolentnější než jiné: např. ke studiu prezenční jednooborové psychologie byli v roce 2012 přijati uchazeči od hranice 81 bodů výše, ke studiu jednooborové historie od 43 bodů výše a ke studiu jednooborové německé filologie od 19 bodů výše (Grigárková, 4. června 2012). Tyto rozdíly si vysvětlujeme především různým zájmem uchazečů o dané obory, resp. poměrem „nabídka/poptávka“; vzhledem k nesystematické varianci těchto hranic napříč obory (a patrně i jednotlivými ročníky přijímacích zkoušek) však není zcela možné tyto rozdíly v našich analýzách statisticky zohlednit. Přehled bakalářských oborů nabízených na FF UP v akad. letech 2011/2012 až 2013/2014 včetně způsobu přijímacího řízení je k nalezení v Příloze 2.

¹⁹ Percentily se vztahují jednak ke všem uchazečům napříč verzemi, jednak k uchazečům v rámci dané verze, slouží však pouze jako orientační přehled a při rozhodování o (ne)přijetí nehrají roli – samotné rozhodnutí se opírá o hrubé skóre, resp. součet bodů z jednotlivých fází přijímacího řízení (SPF, OT, UZ).

²⁰ Zde se v roce 2012 různí informace pro obor „Angličtina se zaměřením na komunitní tlumočení a překlad“, kde Grigárková (4. června 2012) udává poměr 20% SPF + 80% OT/UZ, zatímco ve Zprávě o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2012/2013 (nedat.) nalezneme znovu údaj 30% SPF + 70% OT/UZ.

6 Kritéria kvality testových metod

Veškeré testové, resp. psychodiagnostické metody, které jsou v psychologii používány a na jejichž výsledcích často spočívají rozhodnutí se závažným dopadem pro testovaného, by měly splňovat určité požadavky na kvalitu. Mezi ty základní přitom spadá standardizace a objektivita, férovost, reliabilita a validita (Howitt, & Cramer, 2005). V následujících podkapitolách bude o těchto požadavcích, různých přístupech k jejich konceptualizaci a ověřování pojednáváno jednotlivě, je ovšem třeba mít na mysli, že jsou spolu silně provázány. Nejedná se tedy o jejich dodržování ve smyslu „buď-anebo“; dobré testové metody by je měly v přijatelné míře splňovat všechny.

6.1 Standardizace a objektivita

Termín *standardizace* lze chápat minimálně třemi různými způsoby, jak upozorňuje už Říčan (1977) a později také Urbánek, Denglerová a Širůček (2011):

Zprv se jedná o sjednocení a přesný popis způsobu administrace, vyhodnocení a interpretace testu, resp. dosažených skóre, včetně jednotné vizuální podoby testového sešitu, instrukce, pracovních podmínek atp. Zabezpečení těchto požadavků bývá obvykle nejsnadnější, ačkoliv při přijímacích zkouškách se mohou vyskytnout procesní komplikace: Z kapacitních důvodů učeben na FF UP a množství přihlášených není například možné zajistit, aby všichni uchazeči absolvovali test ve stejných prostorách a v tentýž den a hodinu. Vzhledem k tomu, že o některé obory na FF UP je vyšší zájem než o jiné, může se stát, že část uchazečů o tentýž obor bude k písemnému testu pozvána dopoledne, část odpoledne²¹, případně bude testování probíhat v jiných místnostech. Jakkoliv se FF UP snaží zabezpečit, aby byly prostorové podmínky ve všech případech ekvivalentní, je pravdou, že některé učebny jsou vybaveny např. jinými psacími stoly či židlemi. V případě dopoledního a odpoledního testování je zase možné namítat, že z hlediska přirozených biorytmů jsou uchazeči absolvující test po obědě znevýhodněni, neboť u nich touto dobou klesá pozornost. K tomu je třeba navíc připočíst fakt, že uchazeči v odlišných termínech dostávají různé verze testu, aby se zamezilo zkreslení způsobenému prozrazením otázek. Všechny tyto námitky mohou být pro neúspěšné uchazeče důvodem k odvolání a argumentem „nedostatečné standardizace v prvním slova smyslu“. Bylo by tedy lepší zabezpečit skutečně ekvivalentní, dostatečně velké a vybavené učebny, kde by mohla většina uchazečů absolvovat testy (byť třeba i v několika odlišných verzích)

²¹ Tento postup ovlivní i „složení“ všech uchazečů, kteří absolvují danou verzi testu napříč učebnami.

zároveň. Nutno ovšem podotknout, že otázku standardizace (v prvním významu) nenechává FF UP zcela bez povšimnutí: např. instrukce k vyplňování testů SPF byly doposud zadávány ústně jednotlivými dohlážiteli, od letošního ročníku (2014/2015) se však plánuje opora ve formě Power Pointové prezentace. Dodržení „standardních postupů“ by zároveň mělo zajistit *objektivitu* testu, čili nemožnost zkreslení výsledků jednotlivých uchazečů administrátorem - ať už v průběhu administrace, kdy by zadávající mohl být nevědomky například „napovídat“, při vyhodnocení testu nebo při interpretaci výsledků.

Druhým významem slova „standardizace“ může být převod hrubých skóreů na skóre standardní (jako je z-skór, T-skór, Steny apod.) a tvorba *norem*, čili *normalizace*. Jedná se o proces, kdy je metoda „odzkoušena“ na dostatečně velkém, popř. náležitě stratifikovaném (dle věku, pohlaví, vzdělání atp.) vzorku z cílové populace, pro niž je test určen, a je stanoveno rozložení skóreů v této populaci, jejich průměr a standardní odchylka. Tím je následně možné provést srovnání výkonu jedince s výkonem dalších osob ve stejné populaci. V případě SPF je snahou o toto porovnání použití percentilů, tzn. „seřazení“ uchazečů relativně k výkonu ostatních. Jak již ovšem bylo zmíněno dříve, nejedná se o „pevné“ zakotvení v jediném, předem definovaném a neměnném základním souboru (jako tomu je u běžného procesu normalizace, kdy jsou všechny osoby srovnávány se stále stejným referenčním vzorkem), nýbrž o porovnání s ostatními uchazeči z daného ročníku testování, resp. z dané verze testu. Je tedy možné, že uchazeč se 30 body v jedné verzi SPF bude lepší jak 90% jeho kolegů a dosáhne tak na percentil 90, zatímco uchazeč se 30 body v jiné verzi bude patřit mezi „průměr“ s 50. percentilem (např. proto, že jeho kolegové budou zrovna „zdatnější“). Nejedná se přitom o chybu testu per se (např. odlišné obtížnosti jednotlivých verzí) jako spíš o možnost, že uchazeči absolvující různé verze testu pocházejí z různých cílových sub-populací (např. kdyby uchazeči se zájmem o studium historie vykazovali odlišné schopnosti než uchazeči se zájmem o psychologii), případně že jsou vystaveni odlišným podmínkám testování (viz výše). I v tomto ohledu by tedy bylo dobré stanovit nějaký jednotný postup srovnávání uchazečů, popřípadě zauvažovat o vlivu dalších proměnných na celkové rozhodnutí o (ne)přijetí uchazeče (i v tom smyslu, nespoléhat se na „jediný“ výsledek testu, jak bylo diskutováno dříve v této práci).

Standardizace ve třetím slova smyslu pak zahrnuje kromě tvorby norem i prokázání validity, reliability a objektivy testu, tj. všechna výše i později zmíněná kritéria. Právě o přispění k třetímu typu standardizace u testů SPF na FF UP se snaží tato práce.

6.2 Férovost

Než se pustíme do výkladu pojmu *férovost* v rámci hodnocení kvality psychodiagnostických metod, dovolíme si zamyšlení nad problematikou „férovosti“ v kontextu celého přijímacího řízení. V podmínkách Spojených států amerických jsou totiž přijímací řízení na vysoké škole otázkou prestiže na jedné straně (uchazeči se chtějí dostat na „nejlepší“ školy a univerzity mají zájem být hodnoceny jako „nejlepší“) a férovosti vůči potenciálně znevýhodněným menšinám na straně druhé.

Nad definicí férovosti – stejně jako nad mnoha dalšími v oblasti (nejen) terciárního vzdělávání – však panuje názorový rozpor: V něm jedna část společnosti považuje za „fér“, aby byl přístup k vysokoškolskému vzdělávání garantován příslušníkům různého etnického původu, mateřského jazyka, náboženského vyznání a případně i pohlaví ve stejném poměru, v jakém jsou tito zastoupeni ve většinové populaci, resp. celkové populaci uchazečů.²² Druhá část naopak vidí jako „férové“ hodnotit kvality, schopnosti a úspěchy všech uchazečů bez ohledu na jejich příslušnost ke konkrétnímu etniku, náboženské skupině či pohlaví (Camara, & Kimmel, 2004).

Podobně je férovost coby požadavek na veškeré testové metody (a zvláště pak výkonové testy) i v řadách odborníků vnímána různorodě: Schultz a Whitney (2005) tvrdí, že „férovost“ (*fairness*) – na rozdíl od testového zkreslení (*test bias*) – není psychometrickým, ale sociopolitickým konceptem, a stížnosti na „neférovost“ testu nejčastěji přicházejí od lidí, kteří v něm nedosáhli očekávaných či požadovaných výsledků. Typicky se jedná o skupiny národnostních menšin a lidí s nižším socioekonomickým statutem (SES). Sackett, Borneman a Connelly (2008) ovšem ukazují, že (skutečně poměrně silné, $r = 0,42$) korelace mezi socioekonomickým statutem a výsledkem testu kognitivních schopností jako je SAT neovlivňují vztah mezi SAT a vysokoškolskými známkami (parciální korelace za vyloučení SES, $r = 0,44$). Jde tedy o rozdílná konkrétní dosažená skóre jedinců z různých skupin spíše než o chybu samotného testovacího procesu a jeho rozdílné predikce výkonů pro různé skupiny lidí, tedy o testové zkreslení. Reálné rozdíly mezi kognitivními výkony společenských menšin a většinovou populací skutečně mohou představovat problém ve „férové“ dostupnosti vysokoškolského vzdělání, pakliže jsou testy kognitivních schopností použité coby jediné kritérium pro přijetí, nejsou ovšem udržitelným argumentem pro zpochybnění validity testů testovým zkreslením.

²² V českém prostředí se někdy hovoří o tzv. „pozitivní diskriminaci“, známé především z období po r. 1948 v souvislosti s přijímáním uchazečů o studium na VŠ na základě jejich „kádrového původu“ (Urbášek, 2008).

To může být způsobeno např. přítomností položek, které přímo nesouvisí s testovaným kritériem a zároveň podléhají velkému vlivu faktorů jako je kultura či mateřský jazyk, což způsobuje systematicky odlišné rozložení skóre příslušníků různých menšin. Toto zkreslení lze kontrolovat např. pomocí metody moderované mnohonásobné regresní analýzy, přičemž sledujeme, zda konkrétní skupiny (dělené dle pohlaví, etnické příslušnosti či socioekonomického statutu) vykazují v rovnici, kde závislou proměnnou je predikovaný výsledek kritéria, odlišné beta-koeficienty. Problémem při tomto postupu však může být nestejná velikost srovnávaných skupin (už ze samotné definice slova „menšina“ vyplývá, že zastoupení jejích členů v našem vzorku bude výrazně nižší), která sama o sobě může do rovnice vnést zkreslení (Schultz, & Whitney, 2005). Ke kontrole testového zkreslení lze samozřejmě využít i jiných metod (namátkou: rozpoznávání odlišného fungování položek v rámci teorie odpovědi na položku), jejich výčet a popis ovšem již přesahuje rámeček této práce.

I tvůrci amerických Standardů pro testování ve vzdělávání a psychologii (Joint Committee for Standards et al., 1999) přiznávají nejednoznačnost pojmu „férovost“ a hovoří o čtyřech různých pojetích:

- a) **férovost jako absence testového zkreslení** (viz výše);
- b) **férovost jako stejné zacházení se všemi účastníky v průběhu procesu testování** (souvisí se standardizací nejen metody samotné, ale také průběhu její administrace a vyhodnocení);
- c) **férovost jako stejné rozložení výsledků pro všechny podskupiny** (diskutované výše Schultzem & Whitneym, 2005); a
- d) **férovost jako stejná možnost všech účastníků připravit se na test** (především jde-li o znalostní zkoušku).

S posledním zmíněným pojetím, totiž s možností seznámit se s testovým materiálem, pracuje i Vyhláška č. 343/2002 Sb. o postupu a podmínkách při zveřejnění průběhu přijímacího řízení na vysokých školách. Podle §1 této vyhlášky mají vysoké školy a fakulty povinnost zveřejnit po proběhlém přijímacím řízení mimo jiné „úplné zadání zkušebních otázek či příkladů, které jsou součástí přijímací zkoušky nebo její části, a u otázek s výběrem odpovědi správné řešení“. Tento paragraf, ač z hlediska „férovosti“ vůči budoucím uchazečům ospravedlnitelný, podstatně komplikuje život všem fakultám a oborům, které si přijímací testy tvoří samy, často s omezeným finančním a personálním

rozpočtem. Zahrnuje v sobě totiž nutnost vytvářet každý rok nové otázky (často, vzhledem k počtu uchazečů a kapacitám zkušebních místností, ještě v několika ekvivalentních verzích), jejichž pilotní prověření je vzhledem k riziku prozrazení přinejmenším obtížné, ne-li nemožné. Jejich psychometrické vlastnosti tak mohou být v nejlepším případě odhadovány na základě podobnosti s předchozími položkami stejného typu (např. číselné řady, verbální analogie atp.), empiricky ověřeny jsou však vždy až „post-hoc“. Tím se ovšem komplikuje i celkové vyhodnocení testů, neboť veškeré „vadné“ (špatně diferencující, nedostatečně korelující s celkovým skóre v testu apod.) položky by do skóre jednotlivých uchazečů neměly být započítávány. Pokud navíc v různých verzích testu půjde o jiný počet položek, následná korekce skóre může být minimálně nepřehledná, nemluvě o možné napadnutelnosti „spravedlivosti“ takových testů neúspěšnými uchazeči. I z tohoto důvodu by FF UP a vysoké školy obecně měly usilovat o „neprůstřelnost“ jejich přijímacích testů, i kdyby to znamenalo navýšení rozpočtu na jejich tvorbu.

6.3 Reliabilita

Jak již bylo řečeno v úvodním vymezení pojmů, reliabilita se týká spolehlivosti měření, tzn. zda by titíž lidé dosáhli v (stejných či paralelních formách) testu stejných výsledků, pokud by byli (opakovaně) měřeni za stejných podmínek a zkoumaná vlastnost se u nich nezměnila (Crocker, & Algina, 2008; Urbášek, Denglerová, & Širůček, 2011), a zakládá na specifické variabilitě testových skóre. Podle *klasické testové teorie* (angl. *classical test theory* nebo také *classical true-score model*, zkráceně CTT) je každá naměřená hodnota složená ze dvou komponent, totiž *pravého skóre*, který odráží skutečné schopnosti, dovednosti či vlastnosti jedince, a *chyby měření*. Ta může být dále rozložena na chybu *systematickou* a *nesystematickou*. Systematická chyba se objevuje např. pokud metoda každému zkoušenému konzistentně přidává (nebo ubírá) tři body nebo pokud dotyčný u každé položky, v níž si není jistý odpovědí, zahrhne prostřední možnost – a tato tendence se opakuje při všech zkoušeních. Systematickou chybu, pakliže je odhalena, je možné celkem snadno statisticky ošetřit, a v psychometrii nepůsobí tolik obtíží jako chyba *nesystematická*. Ta vzniká například hádáním, odlišným naladěním jednotlivců v testové situaci, chybami při zadávání nebo jinými čistě náhodnými vlivy a není možné přesně určit, v jaké směru a jaké míře jednotlivé výsledky zkresluje (Crocker, & Algina, 2008; Říčan, 1977).

Vzhledem k tomu, že momentální stav se u lidí velmi rychle mění, je v psychologii velmi nepravděpodobné, že by tentýž člověk vždy dosáhl naprosto identických skóre. Lze tedy počítat s tím, že výsledek každého testu nebo testování nebude přesně odpovídat pravému skóru jedince, ale bude se pohybovat někde kolem této hodnoty. Celkový rozptyl skóre všech testovaných je tedy možné označit za součet *rozptylu pravých skóre* a *chybového rozptylu*. Pakliže má být metoda reliabilní, usilujeme o to, aby právě její náhodný chybový rozptyl byl pokud možno co nejnižší (Crocker, & Algina, 2008).

V rámci CTT existují v zásadě tři způsoby, jak reliabilitu ověřovat; hovoříme proto někdy o třech *druzích* reliability: na základě opakovaného předložení téhož testu stejným probandům ve dvou různých časových bodech (*test-retestová reliabilita* nebo též *stabilita v čase*), srovnáním výkonu stejných osob ve dvou různých formách testu předložených současně (*reliabilita paralelních forem* či také *ekvivalence*) nebo porovnáním míry správných a špatných odpovědí každého zkoušeného v různých částech testu, případně na úrovni jednotlivých položek (*reliabilita vnitřní konzistence* nebo také *split-half reliabilita*) (Urbánek, Denglerová, & Širůček, 2011). Veškerá tato srovnání jsou založena na korelačních koeficientech mezi pozorovanými skóre v testu (do nichž se promítnou jednak pravé skóre, jednak právě velikost chyby měření), přičemž je třeba mít na paměti, že vypočtené koeficienty jsou spíše odhady a velmi závisí na konkrétním vzorku uchazečů, který právě testujeme (Crocker, & Algina, 2008). Schultz a Whitney (2005) přitom pro všechny druhy uvádějí, že v psychologii přijatelná spodní hranice korelačního koeficientu, aby mohla být metoda považována za reliabilní, je $r = 0,7$ (srov. Crocker, & Algina, 2008).

Jelikož je v případě SPF každý test administrován uchazečům pouze jednou, dovolíme si pro zkrácení délky této práce vynechat popis výpočtu test-retestové reliability; při znalosti principu korelace je ovšem poměrně nasnadě. Naproti tomu v jednotlivých ročnících SPF byly vytvářeny i vzájemně paralelní verze²³, ačkoliv nebyly předkládány týmž uchazečům; pokusíme se tedy stručně shrnout standardní metody ověřování reliability paralelních forem a hrubě nastínit námi aplikovaný postup. Ten se bude opírat především o reliabilitu vnitřní konzistence, proto této metodě věnujeme více prostoru.

²³ Viz kapitola o Testu předpokladů ke studiu na FF UP a dále kapitola Reliabilita paralelních forem.

6.3.1 Reliabilita paralelních forem

Jak uvádějí Crocker a Algina (2008), dvě verze testu je možné označit za paralelní, pokud je pravý skór každého testovaného v obou verzích stejný a zároveň mají obě verze shodné chybové rozptyly; stejně tak by se neměly lišit průměry a směrodatné odchylky jednotlivých testů a skóry uchazečů z obou verzí by společně měly vysoce korelovat. Urbánek, Denglerová a Širůček (2011) hovoří o různých pojetích paralelity, jednak jako srovnatelnosti obsahu, instrukce a způsobu administrace, jednak o ekvivalenci získaných skórů z hlediska interpretace a v neposlední řadě jako o shodě parametrů (průměrů, směrodatných odchylek a korelací s jinými psychodiagn. metodami) různých forem testu. Empirické zhodnocení reliability paralelních forem poté dle autorů probíhá porovnáním průměrů a rozptylů jednotlivých položek a vzájemné kovariance „mezi všemi položkami v obou formách“ (Urbánek, Denglerová, & Širůček, 2011, 102). Tato poněkud víceznačná formulace se ovšem týká situace, kdy srovnáváme a korelujeme skóry *týchž* probandů v obou testech, a jejich výkony by se tedy neměly napříč testy příliš lišit.

V případě SPF ovšem každý uchazeč absolvuje pouze jednu verzi testu. Byla však vyvinuta snaha, aby si jednak:

- a) všechny testové verze byly formálně podobné, tzn. aby obsahovaly stejné instrukce, formát položek a jejich vyplňování, stejně jako způsob vyhodnocení, a také stejné subtesty (verbální myšlení, kritické myšlení, logicko-analytické myšlení, všeobecný přehled), ačkoliv tyto jsou v sudých a lichých verzích uvedeny v odlišném pořadí²⁴;
- b) všechny testové verze byly podobné obsahově, tzn. obsahovaly stejný počet otázek v každém subtestu (viz Tabulka 2), tyto otázky cílily na stejné či podobné schopnosti a vědomosti (vč. subtestu všeobecného přehledu) a byly pokud možno ekvivalentně náročné,

přičemž zároveň existovaly přímo páry verzí (lichá-sudá), které sestávaly z identických položek, pouze v jiném pořadí a/nebo s jiným pořadím správných odpovědí²⁵. Z tohoto hlediska bychom jednotlivé verze SPF mohli považovat za (alespoň formálně) srovnatelné.

²⁴ Viz níže.

²⁵ Tyto páry byly vždy administrovány současně v témže termínu napříč učebnami, tzn. v každém termínu dostala polovina uchazečů ve všech učebnách lichou verzi a polovina odpovídající sudou verzi, přičemž v termínu následujícím se postup opakoval s „novými“ (lichými) verzemi.

Jak již bylo diskutováno v kapitole o standardizaci, existuje možnost, že uchazeči absolvující různé (liché) verze testu, které byly administrovány pokaždé v odlišných termínech, pocházejí z různých sub-populací, případně že jsou vystaveni odlišným podmínkám testování. Srovnání průměrů a rozptylů jednotlivých (nepárových) verzí testu, které bychom normálně prováděli ke zjištění reliability paralelních forem, pokud by každý uchazeč absolvoval všechny verze, tedy v našem případě může odrážet několik druhů rozdílů: jednak rozdílnost verzí jako takových, jednak potenciální rozdílnost pravých skóre mezi uchazeči v jednotlivých skupinách a jednak chyby potenciálně způsobené odlišným časem začátku administrace (a tedy možnou únavou respondentů). Právě proto, že žádný uchazeč neabsolvuje více než jednu verzi, bychom tedy při použití klasických technik posuzování reliability paralelních forem (lichých verzí) de facto srovnávali jablka a hrušky, resp. bychom nebyli schopni odlišit, jaká část chybového rozptylu je připisatelná kterému (potenciálnímu) zdroji chyb – faktickým schopnostem uchazečů, únavě, nebo verzi SPF – a tedy jasně rozhodnout, zda jsou si všechny verze striktně empiricky paralelní²⁶.

Na druhou stranu (intuitivně) předpokládáme, že minimálně páry lichých a sudých verzí, které obsahují identické položky a byly administrovány v tutéž dobu („stejně skupině“), by měly vykazovat podobné parametry a hodnoty vnitřní konzistence. Pokud by tomu tak nebylo, můžeme uvažovat, že i „prohození“ pořadí otázek či správných odpovědí může významně ovlivnit výsledné skóre respondentů. Omezíme se tedy v naší práci, pro první „hrubý odhad“, na srovnání těchto párových verzí v každém ročníku, a to konkrétně na porovnání průměrných celkových skóre dosažených uchazeči v každé verzi (pomocí dvouvýběrového t-testu), rozptylů těchto skóre (F-test) a porovnání korelačních koeficientů vnitřní konzistence jednotlivých verzí (z-test), s tím, že uznáváme, že tento postup není zcela standardní a nedozvíme se z něj informace o vzájemné paralelitě *všech* verzí.

6.3.2 Reliabilita vnitřní konzistence

Koncept vnitřní konzistence se zakládá na podobné logice jako paralelní verze testů, totiž že má-li test měřit stejnou vlastnost, pak by tentýž člověk měl dosahovat stejného pravého skóre nezávisle na tom, jakou část všech potenciálních položek zrovna zodpoví (tzn. jakou z verzí testu dostane). Pokud bychom tento předpoklad aplikovali na jediný test, pak by např. výsledek dotyčného v první polovině a ve druhé polovině testových položek

²⁶ I kdybychom např. na základě jednofaktorové analýzy variance nenašli rozdíly mezi průměry testů, stejně nebudeme schopni určit, jakou kombinací vlivů tato „podobnost“ vznikla.

měl být ekvivalentní a jednotlivé položky by měly navzájem korelovat (Říčan, 1977; Urbánek, Denglerová, & Širůček, 2011).

Jedním ze způsobů ověřování vnitřní konzistence je tzv. metoda *split-half*, čili rozložení testu na dvě (stejně dlouhé) poloviny a výpočet korelací skóre uchazečů v první a druhé půlce. K tomu se využívá *Spearman-Brownova vzorce* uvedeného na *Obrázku 1*, kde r_{AB} představuje korelaci mezi první a druhou polovinou testu a $r_{xx'}$ odhad celkové vnitřní konzistence.

$$r_{xx'} = \frac{2r_{AB}}{1 + r_{AB}}$$

Obr. 1: Spearman-Brownův vzorec (Crocker, & Algina, 2008)

Zobecnění Spearman-Brownova vzorce také ukazuje, že reliabilitu testu je možné „uměle“ zvyšovat přidáváním položek (měřících stejnou vlastnost); odhad vnitřní konzistence tedy přímo závisí na délce testu (Říčan, 1977). Zároveň se při použití metody *split-half* vystavujeme problému s nejednoznačným odhadem, neboť test může být rozdělen na poloviny mnoha různými způsoby a korelace takto sestavených polovin nemusí být zcela identické (Crocker, & Algina, 2008). I proto je možné a v mnoha případech žádoucí zvolit jednu z metod založených na kovarianci mezi položkami, jako je např. *Cronbachův koeficient alfa* nebo *Kuder-Richardsonův vzorec*.

Výpočet *Cronbachova koeficientu alfa* je založen na známém rozptylu testových skóre a kovariační matici všech jednotlivých položek a jeho vzorec zachycuje *Obrázek 2*, kde K je počet položek v testu, σ_i^2 představuje rozptyl jednotlivé položky i a σ_x^2 je celkový rozptyl skóre v testu.

$$\alpha = \frac{K}{K - 1} \left(1 - \frac{\sum_{i=1}^K \sigma_i^2}{\sigma_x^2} \right)$$

Obr. 2: Cronbachova alfa (Crocker, & Algina, 2008)

Alfu je možné použít jak pro dichotomně skórované položky (0/1 čili špatně/správně, jako je tomu u SPF), tak pro položky polytomické (např. odpověďové škály). Hodnota tohoto koeficientu nám ukáže nejnižší mez očekávané reliability, neboť v reálném testování nelze očekávat, že všechny „součásti“ testu (poloviny, subtesty, příp. jednotlivé položky) budou dokonale paralelní, a tudíž teoreticky předpokládaná reliabilita bude vyšší než reliabilita „skutečně pozorovaná“ (Říčan, 1977).

Další možností odhadu vnitřní konzistence testu je použití tzv. *Kuder-Richardsonova vzorce 20*, který je vhodný konkrétně pro dichotomické položky. *Obrázek 3* ukazuje výpočet KR-20, přičemž K je opět počet položek, σ_x^2 je celkový rozptyl testu a $p_i q_i$ je zjednodušený výpočet rozptylu položky, kde p_i představuje podíl osob, které mají danou položku i správně, a q_i podíl osob, které mají položku i špatně (platí $q_i = 1 - p_i$).

$$r = \frac{K}{K - 1} \left[1 - \frac{\sum_{i=1}^K p_i q_i}{\sigma_x^2} \right]$$

Obr. 3: *Kuder-Richardson 20* (Crocker, & Algina, 2008)

Porovnáním obou vzorců (KR-20 a Cronbachovy alfy) můžeme zjistit, že u dichotomických položek se jedná prakticky o shodný výpočet, ačkoliv Urbánek, Denglerová a Širůček (2011) poukazují na fakt, že někteří autoři vyzdvihují použití KR-20 v případě, kdy se jedná o časově omezený výkonový test, v němž není možné zodpovědět všechny položky. V takové situaci může totiž být výpočet vnitřní konzistence testu zkreslený, což by KR-20 měla údajně řešit. Urbánek, Denglerová a Širůček (2011, 108) však upozorňují, že tuto schopnost KR-20 nemá, i když pro odhad vnitřní konzistence testů s dichotomně skórovanými položkami je vhodnější než Cronbachův koeficient alfa, neboť ten je „zodpovědně použitelný pouze pro položky, jejichž skóry sledují normální rozdělení“. K výpočtu reliability coby vnitřní konzistence testů tedy v případě SPF použijeme právě KR-20.²⁷

Abychom vůbec mohli uvažovat, že výkon jednotlivce bude napříč položkami konzistentní, je potřeba, aby tyto položky (testové otázky) byly *homogenní*, tzn. měřily tytéž schopnosti, vlastnosti či dovednosti, ale také aby se v nich nevyskytovaly chyby v zadání apod. (Crocker, & Algina, 2008). Právě homogenita položek a měřené schopnosti jsou hlavním sporným bodem v konstruktové validitě testů studijních předpokladů (zde se tedy validita a reliabilita prolíná): Pokud se náš test, podobně jako v případě SPF, skládá z různých částí (verbální myšlení, logicko-analytické myšlení apod.) obsahujících různé typy položek a tedy potenciálně cílících na poněkud odlišné dovednosti, je vůbec možné dosáhnout vysoké reliability ve smyslu vnitřní konzistence testu? A pokud ano, nenahrává toto zjištění argumentu jediné „obecné kognitivní schopnosti“ ve smyslu g-faktoru?²⁸

²⁷ Se zkrácením způsobeným časovým limitem u SPF nepočítáme, neboť 60 minut na 50 položek se z pozorování uchazečů v uplynulých ročnících ukázalo jako dostatečné množství času.

²⁸ Crocker a Algina (2008) upozorňují, že vysoký koeficient alfa neznamená automaticky unidimenzionalitu testu, jako spíš možnost, že různé faktory, na které jednotlivé položky cílí, spolu do určité míry korelují.

Otázka dimenzionality testů studijních předpokladů, resp. testu SPF konkrétně, bude řešena níže v kapitolách o validitě a položkové analýze; nyní se ještě zaměříme na některá omezení při výpočtech reliability podle klasické testové teorie.

Jedním z těchto omezení je fakt, že výpočet vnitřní konzistence testu přímo závisí za počtu položek v testu; čím vyšší počet položek, tím „lépe“ je teoreticky postihnuta dimenze, kterou zamýšlíme měřit, a tím větší je tedy spolehlivost odhadu pravých skóreů testovaných. Zároveň je potřeba, aby zkoušení měli na zodpovězení všech položek dost času, jinak je koeficient vnitřní konzistence zkreslený. V neposlední řadě pak výsledek ovlivní i konkrétní zkoumaný vzorek, na němž byly výpočty prováděny (neboť veškeré korelace se zakládají na konkrétních skórech konkrétních jedinců), a jeho homogenita (viz omezení variability v kapitole 1) – pokud není soubor uchazečů dostatečně variabilní co do pravých skóreů, většina testové variance bude připsána chybovému rozptylu (Říčan, 1977).

Ačkoliv k odhadům reliability existují i odlišné přístupy (např. teorie zobecnitelnosti), jež se snaží některá z výše uvedených omezení řešit, jejich výpočty by v případě SPF byly opět nevhodné či neúměrně komplikované zamýšlenému využití výsledků. Jelikož v současnosti nemáme dostatek „prověřených“, resp. vůbec „prověřitelných“ položek, které bychom mohli v testech SPF využít opakovaně, aniž bychom se vystavovali přílišnému riziku zkreslení skóreů jednotlivých uchazečů právě známostí otázek předem²⁹, stačí nám zatím k základnímu posouzení psychometrických vlastností odhad vázaný na konkrétní uchazeče, kteří test absolvovali. I tak je ovšem žádoucí, aby tento vzorek co nejlépe odpovídal (variabilní) populaci, pro kterou test zamýšlíme. Měli bychom si tedy ujasnit, koho vlastně za tuto populaci považujeme: Na jedné straně můžeme „cílovou populací“ označit všechny uchazeče o studium na FF UP, přičemž v každém ročníku se k přijímacím zkouškám přihlásí právě tito zájemci a test SPF absolvují všichni (a vzorek tedy bude prakticky perfektně reprezentativní). Na straně druhé můžeme uvažovat o „potenciálních uchazečích“, resp. „zbytku populace (ČR)“, která se na vysokou školu z různých důvodů nehlásí, ovšem jejíž předpoklady ke studiu by teoreticky také bylo možné testem zjišťovat. V tomto případě pak lze pochybovat, zda uchazeči o studium na VŠ (nebo FF UP konkrétně) náhodou již předběžně nevykazují vyšší studijní předpoklady než „zbytek populace“, což je také vedlo k rozhodnutí se na vysokou školu

²⁹ Bavíme se o stovkách či tisícovkách položek, neboť při takovém množství předpokládáme, že není v běžných lidských silách naučit se z paměti odpovědi na všechny položky, i kdyby měl uchazeč předem k dispozici jejich seznam, znění či vyhodnocení. V takové situaci by bylo v našem zájmu znát přesné, na uchazečích nezávislé parametry jednotlivých položek, a do jednotlivých verzí vybírat pouze ty kvalitní.

hlásit. Na tyto otázky, podobně jako na mnohé další konceptuálně provázané se studijními předpoklady, ovšem prozatím neznáme jednoznačnou odpověď, což opět omezuje interpretační možnosti analýz v této práci.

6.4 Validita

Schultz a Whitney (2005) o „platnosti metody“ říkají, že při jejím zjišťování nejde ani tak o to, zda je platný samotný test, jako spíš zda jsou platné závěry a interpretace, které na základě skóre dosažených v tomto testu budeme činit. Tím poukazují i na více úhlů pohledu, ze kterých lze validitu nahlížet: Ve starších učebnicích psychometrie (např. Říčan, 1977) nalezneme dělení validity na pojmovou a empirickou, resp. výčet různých typů validit. Současné publikace hovoří spíš o různých zdrojích důkazů o validitě, a to *obsahových*, *empirických/kritériových* a *konstruktových* (Schultz, & Whitney, 2005; Urbánek, Denglerová, & Širůček, 2011). Obě pojetí se do značné míry překrývají a některé konkrétní postupy odhadu validity je těžké zařadit pouze do jedné „škatulky“, pro přehlednost se ovšem opět budeme jednotlivým zdrojům a způsobům jejich získávání věnovat odděleně.

6.4.1 Obsahové zdroje důkazů o validitě

Pod obsahové zdroje validity řadí Urbánek, Denglerová a Širůček (2011) jednak obhajitelnost obsahu metody (tzn. volbu a vlastnosti položek) vzhledem k zamýšlenému účelu testování, jednak i charakteristiky testové situace a chování probandů v ní. Říčan (1977) v této souvislosti uvažuje o relevanci „zkouškové situace“ vzhledem k budoucí náplni práce a výkonu v ní (který má být testem předpovězen), resp. o tom, zda se proband bude ve zkouškové situaci chovat stejně jako v budoucnu při měření kritéria. Zvažuje totiž, že pokud na výsledku zkouškové situace velmi závisí další osud jejich účastníků (např. přijetí ke studiu na VŠ), mohou se snažit pozitivně zkreslovat své výsledky, což platí zvláště pro osobnostní či postojoyé dotazníky (z nichž je na první pohled vidět, na co cílí, a lze odhadnout „sociálně žádoucí“ odpověď), ale i pro testy znalostní (zvláště není-li špatná odpověď penalizována mohou uchazeči „alespoň hádat“, neznají-li správnou odpověď). Naproti tomu účastníci testování opakovaně (třeba v průběhu VŠ studia) už takový tlak na výkon pociťovat nemusí, a jejich skóre v daných testech se tedy mohou podstatně lišit. Urbánek, Denglerová a Širůček (2011) nazývají schopnost probanda – laika v oboru psychologie „otipovat“, co metoda nejspíše měří, *zjevnou* validitou metody, a poukazují na její přednosti a nedostatky: Pokud dotyčný vidí „smysl“ testu, může to zvýšit jeho motivaci

při vyplňování, na druhou stranu se však vystavujeme výše zmíněným rizikům zkreslení odpovědí.

Další možností posouzení obsahové validity metody je expertní odhad, nebo také *výběrová* validita testu (Schultz, & Whitney, 2005; Urbánek, Denglerová & Širůček, 2011). Týká se samotného výběru položek vzhledem k vlastnostem, schopnostem, dovednostem nebo znalostem, které má test měřit, a lze ji nechat zhodnotit právě panelem expertů na tuto (zamýšlenou) oblast. Obvyklý postup spočívá nejprve v podrobné definici cílového kritéria (např. studijních předpokladů) a jeho rozdělení na jednotlivé dimenze či oblasti, které by test měl pokrýt. Následně se zhodnotí teoretické proporce těchto oblastí v celém konstrukt (např. nakoľik významné jsou faktické znalosti, kritické myšlení či prostorová představivost) a počet položek v testu, jenž by měl těmto proporcím odpovídat. Experti také mohou posoudit, nakoľik zvolené položky skutečně pokrývají zamýšlenou oblast a zda například není nějaká důležitá dovednost opomenuta (Urbánek, Denglerová, & Širůček, 2011). Schultz a Whitney (2005) hovoří při tomto způsobu posuzování také o použití tzv. *content validity ratio*, což by se dalo volně přeložit jako „poměr obsahové validity“ (Obr. 4). Jedná se o index „vhodnosti“ pro každou položku, sestavený na základě hodnocení několika posuzovatelů jako poměr počtu těch, kteří položku považují za důležitou (n), vůči celkovému počtu hodnotitelů (N). Z průměru těchto indexů lze pak sestavit celkový index obsahové validity daného testu.

$$CVR = \frac{n - \left(\frac{N}{2}\right)}{\frac{N}{2}}$$

Obr. 4: *Content validity ratio* (dle Schultz, & Whitney, 2005)

Vzhledem k problematickému konstrukt studijních předpokladů a procesu tvorby testů SPF na FF UP představuje zhodnocení obsahové validity testů SPF i za použití tohoto postupu metodologický „oříšek“. Snaha o posouzení expertním panelem probíhá de facto v průběhu samotné tvorby jednotlivých ročníků testu, na níž se podílejí zástupci z řad různých kateder FF UP (především pak lingvisté, ale i ekonomové, historici a psychologové), kteří společně vybírají a hodnotí navrhované položky pro každý ročník a verze testu. Je tedy třeba, aby se tito lidé shodli na položkách, které budou do jednotlivých verzí zařazeny. Problémem je možné zkreslení vzniklé faktem, že tito posuzovatelé jsou zároveň *tvůrci* testu, a mohou tak být příliš „ponořeni“ do svého pohledu na věc.

Objektivní zhodnocení nezávislým odborníkem by tedy mohlo být vhodnější, je však též otázkou financí a navýšení rizika prozrazení aktuálního znění testů.

Výše uvedeného dělení konstruktů na různé oblasti se týká také *faktorová* validita, která, jak lze dovodit z názvu, se zakládá na metodě faktorové analýzy odpovědí velkého množství respondentů na jednotlivé položky ověřovaného testu. Faktorová analýza vychází z předpokladu, že v pozadí velkého množství položek stojí jen několik málo společných faktorů, které ovlivňují odpověď testovaného na danou položku (Crocker, & Algina, 2008). Těmto faktorům se také říká *dimenze* a hovoří se proto i o *dimenzionalitě* testů: Testy měřící jedinou schopnost (např. mentální rotace objektů) lze označit jako unidimenzionální, testy cílící naráz na více aspektů (byť jediného konstruktů) jsou považovány za multidimenzionální (Říčan, 1977). Urbánek, Denglerová a Širůček (2011) uvádějí, že unidimenzionální testy by měly v (prvním) společném faktoru vykazovat náboj alespoň 0,3.

O problematice dimenzionality testů studijních předpokladů jsme se zmiňovali již v kapitole o reliabilitě, kde bylo řečeno, že vnitřní konzistence nemusí nutně znamenat unidimenzionalitu testu. Hlavní důvody, proč se touto otázkou (znovu) zabýváme, jsou v zásadě dva. První lze nazvat „teoretickým“ a týká se předpokladů ke studiu jako konstruktů, který byl diskutován již v kapitole 4.1.1: Pokud bychom v rámci faktorové analýzy odhalili faktory korespondující s jednotlivými subtesty, mohli bychom toto zjištění považovat za argument pro existenci *různých* studijních předpokladů, resp. různých (a poněkud odlišných) schopností, dovedností a znalostí, které se SPF pokouší testovat. Na druhé straně pokud bychom byli schopni identifikovat pouze jediný společný faktor, můžeme se dohadovat, že veškeré „dílní“ schopnosti, znalosti a dovednosti jsou pouze různými projevy jediné „obecné kognitivní schopnosti“, jak o ní hovoří např. Carroll (1993) či Gottfredsonová (2004).

Druhý důvod pro diskuzi dimenzionality je „praktičtějšího“ rázu, neboť některé teorie zabývající se kvalitou testových položek (např. teorie odpovědi na položku) mají při své aplikaci na dimenzionalitu testu pevné požadavky (viz Urbánek, Denglerová, & Širůček, 2011). I proto se v naší práci nejprve zaměříme na analýzy za pomoci klasické testové teorie (jež tak striktní podmínky na dimenzionalitu nemá) a faktorové analýzy, abychom byli následně schopni odhadnout možnosti využití dalších přístupů při hodnocení kvality testů SPF a jednotlivých položek.

6.4.2 Empirické/Kritériové zdroje důkazů o validitě

Jak je z názvu patrné, vypovídají empirické (či také kritériové) zdroje validity o vztahu mezi výsledkem (skórem) zkoumaného testu a nějakým vnějším kritériem (indikátorem zkoumané vlastnosti, schopnosti atp.). Podobně jako u validity obsahové existuje i zde více možností, jak kritériovou validitu ověřovat: V první řadě můžeme hovořit o validitě *souběžné*, kdy se snažíme zjistit míru nějakého jevu či vlastnosti, která se u respondenta vyskytuje již za průběhu testování (např. přítomnost vývojové poruchy či rizikového chování), nebo o validitě *prediktivní*, kdy se snažíme předpovídat výskyt/projev nějaké vlastnosti, schopnosti, či situace v budoucnosti. Korelovat lze přitom proměnné zjištěné na základě pozorování či výsledků jiných testů, měření a vyšetření, musíme však mít na paměti, že ani tato nemusejí být stoprocentně „platná“ či „spolehlivá“ (viz níže) (Říčan, 1977; Urbánek, Denglerová, & Širůček, 2011).

Zároveň nemusíme vždy zjišťovat míru „shody“ mezi testem a kritériem, ale i míru „neshody“ – například pokud chceme rozlišit dva související či podobné, ne však úplně stejné konstrukty (např. krátkodobou paměť od pozornosti, nebo, jak bychom mohli chtít v našem případě, obecnou inteligenci od studijních schopností). Pokud skóry probandů v testu paměti příliš nekorelují s jejich skórem v testu pozornosti, můžeme říct, že paměťový test (ale de facto i test pozornosti) vykazuje *diferenciální* validitu (Urbánek, Denglerová, & Širůček, 2011).

Říčan (1977, 80) vhodně upozorňuje, že „kolik kritérií, tolik validit“. I proto je podstatná definice konstruktu, který chceme měřit (např. studijní schopnosti) a způsobu jeho ověřování (v našem případě operacionalizace „akademického úspěchu“). Zjištěná validita našeho testu totiž závisí i na kvalitě kritéria, resp. metodě jeho zachycení a kvantifikace (např. spolehlivosti a konzistenci školního známkování napříč obory a učiteli). Existují sice postupy pro korekci „chybovosti kritéria“, v praxi ovšem mají jeden zásadní nedostatek: Tyto korekce počítají se znalostí reliability kritéria, v případě námi zamýšlených analýz tedy se znalostí reliability známkového průměru v prvním, popř. druhém roce studia na FF UP. Jelikož nemáme možnosti, jak tuto reliabilitu zjistit (nelze studenty oznamkovat znovu), není pro nás možné tyto metody použít.

Stejně jako při ověřování reliability testu je i pro zjištěnou validitu důležitá informace o populaci či vzorku, pro které údaje o validitě platí (Říčan, 1977). I zde se potýkáme s problémem omezení variability, a to jednak na prediktoru, jednak na kritériu:

Pokud budou námi vybraní uchazeči dosahovat známkového průměru 1.0-2.0, může to samozřejmě být znakem, že dotyční se učí dobře (a byli tedy po právu vybráni), nevíme však, jak by se učili či jakých známek by dosáhli ti, které jsme nepřijali – podobně jako to nevíme o zbytku populace, která se nehlásila vůbec. Potíží je i variance samotného kritéria (neboť kde není variabilita, není nejen reliabilita, ale ani validita), která v případě VŠ známek není moc vysoká – už samotná škála známek 1-3, potažmo 4 (A-E, resp. F) je dost úzká, i kdybychom počítali známkové průměry a vážili je kreditovou hodnotou předmětů. Do hry přitom vstupuje ještě potenciální různorodost známkování napříč obory a riziko určité „mírnosti“ některých učitelů, tzn. riziko, že počet „lepších“ známek bude převažovat počet „horších“, bez ohledu na faktický výkon či schopnosti studenta (Schultz, & Whitney, 2005; Urbánek, Denglerová, & Širůček, 2011).

Variabilitu, resp. reliabilitu školních známek (kritéria) nemáme příliš možnost ovlivnit, můžeme se však pokusit o méně zkreslený odhad prediktivní validity na základě korekce omezení variability našeho prediktoru (skóru v testu SPF). Jak píše Schmidt, Oh a Le (2006), omezení variability lze dělit na *přímé*, kdy je zkoumaný soubor vybrán právě na základě výsledků testu, jehož validita (vůči nějakému kritériu) je předmětem ověřování, a *nepřímé*, kdy do výběru souboru vstupují ještě další proměnné. V případě přijímacího řízení na FF UP, pokud jsou uchazeči vybíráni na základě kombinace skóre z SPF, oborového testu a/nebo ústní zkoušky, uvažovat o nepřímém omezení variability, autoři ovšem upozorňují, že pokud jsou v tzv. kompozitním skóru (celkovém součtu bodů u přijímacích zkoušek) pouze dvě nebo tři proměnné, je korekce na základě přímého omezení variability přesnější. Výpočet této korekce je znázorněn na *Obrázku 5*, kde R_{XY} je korelace mezi testovým skórem a kritériem po provedení korekce, r_{xy} je tatáž korelace před korekcí, S_X je rozptyl skóre „neomezené“ skupiny (v našem případě všech účastníků SPF) a s_x je rozptyl skóre „omezené“ skupiny (přijatých studentů 1. ročníku).

$$R_{XY} = \frac{r_{xy}(S_X/s_x)}{\sqrt{1 - r_{xy}^2 + r_{xy}^2(S_X/s_x)^2}}$$

Obr. 5: Korekce přímého omezení variability (dle Schultz, & Whitney, 2005)

I za použití této korekce je však dobré mít na paměti, že při nízké variabilitě prediktoru i kritéria případná nízká hodnota vzájemného korelačního koeficientu nemusí

nutně znamenat, že test je nevalidní³⁰, jako spíš že k jeho validizaci byly nedostatečně příznivé podmínky. Zjištěný korelační koeficient by se tak podobně jako u reliability dal považovat za odhad spodní hranice skutečné validity testu (Říčan, 1977).

Vedle (korigovaných i nekorigovaných) korelací je dalším vhodným statistickým postupem ke zjišťování prediktivní validity regresní analýza, která umožňuje využít k předpovědi hodnot kritéria i kombinaci několika prediktorů (jak metrických, tak – v případě specializovaných metod v rámci tzv. *general linear model*, který je rozšířením konceptu mnohonásobné regrese, kategoriálních). To je u ověřování prediktivní validity testů SPF velmi výhodné a žádoucí, neboť můžeme sledovat i přínos proměnných jako je známkový průměr z maturitního vysvědčení, věk uchazeče apod. Tím narážíme na další typ empirické validity, totiž na validitu *inkrementální* čili přírůstkovou. Ta se týká míry, do jaké další proměnné v rovnici přinesou nové informace, resp. zlepší předpověď kritéria. V pojmech mnohonásobné regresní analýzy se jedná o přírůstek R^2 čili *koeficientu determinace*, který vyjadřuje variabilitu kritéria vysvětlenou použitými prediktory³¹ (Eid, Gollwitzer, & Schmitt, 2010; Urbánek, Denglerová, & Širůček, 2011).

Na rozdíl od jednoduché lineární regrese vyjadřují beta-koeficienty v mnohonásobné regresní analýze vliv dané proměnné nad rámec ostatních proměnných. To s sebou přináší dvě podstatné implikace: Zaprvé závisí velikost konkrétního beta-koeficientu mj. na pořadí, v jakém proměnné do rovnice zadáváme. Pokud bychom tedy zadávali prediktory postupně, bude se jejich beta-koeficient lišit v závislosti na tom, jaké proměnné budou do rovnice zadány dřív. Zadruhé můžeme proměnné do rovnice zadávat jednak v původní metrice, jednak standardizované (do podoby z-skóru). První postup se hodí, chceme-li následně srovnávat beta-koeficienty např. pro různé skupiny probandů (typicky muži-ženy, viz kapitola o férovosti; můžeme se ale také pokusit srovnat přínos testu SPF napříč různými obory), druhý je vhodný pro porovnání samotných proměnných (ve smyslu srovnání týchž proměnných měřených jinými nástroji, např. v rámci různých studií). Celkový koeficient determinace (po zadání všech proměnných) by se při použití kteréhokoliv z výše uvedených postupů měnit neměl, statistická významnost jednotlivých

³⁰ Helmstadter (1964, in Urbánek, Denglerová, & Širůček, 2011, s. 130) tvrdí, že se „maximální možná validita rovná odmocnině z reliability“. Schultz a Whitney (2005) upozorňují na zjištění, že korelační koeficient u validity zřídka přesahuje hodnotu $r = 0,5$, a navrhují interpretaci, kdy korelace $r = 0,1$ lze považovat za slabé, $r = 0,3$ za středně silné a $r = 0,5$ za silné.

³¹ Jedná se v podstatě o druhou mocninu mnohonásobné korelace mezi nezávislými proměnnými a závislou proměnnou.

přírůstků R^2 by však mohla s ohledem na pořadí zadávání proměnných kolísat (Eid, Gollwitzer, & Schmitt, 2010).

Jak jsme právě naznačili, při více než dvou nezávislých proměnných existuje více možností zadávání prediktorů do rovnice a tedy porovnávání jejich relativního přínosu: Jedna z nich je zadání všech proměnných do modelu najednou a posuzovat jejich vliv na predikci z hlediska jejich beta-koeficientů; tento přístup však může být problematický, pokud se metrika všech proměnných výrazně liší. Z takto zadaného modelu můžeme ovšem „ubírat“ jednotlivé prediktory a sledovat, o kolik se R^2 změní. Podobně jako u postupného přidávání prediktorů do modelu je však potřeba mít rozmyšlené, na základě jakých předpokladů či jakého pořadí budeme proměnné ubírat/přidávat. Jedna možnost je předchozí teoretická úvaha o očekávané přínosnosti proměnných (jednotlivě, nebo v rámci určitých „bloků proměnných“ zadávaných do modelu naráz), druhá se zakládá na empiricky zjištěných korelacích. „Správný postup“ přitom obecně neexistuje a vždy závisí na konkrétní výzkumné otázce (Eid, Gollwitzer, & Schmitt, 2010).

Ačkoliv není přesně stanovena hranice, od níž je přírůstkovou validitu „nového“ prediktoru možné považovat za významnou³², je dobré zjistit, jak si námi ověřovaný test v předpovědi stojí vedle dalších proměnných a někdy také, zda se ho s ohledem na časovou a další zátěž vůbec vyplatí používat (Říčan, 1977).

6.4.3 Konstrukové zdroje důkazů o validitě

Konstrukovou validitou je myšlen fakt, zda test skutečně zjišťuje konstrukt, o němž tvrdí, že je předmětem jeho měření. Ověření tohoto typu validity je asi nejsložitější, neboť se netýká jen samotného obsahu testu a jeho empirické validity, ale celé teorie v pozadí (údajně) zjišťovaných schopností, vlastností, dovedností či znalostí (Urbánek, Denglerová, & Širůček, 2011) – v našem případě tedy teorie o tom, co jsou „předpoklady ke studiu“ (viz dřívější kapitoly).

Proto také neexistuje jediný či dokonce jednotný postup, číslo nebo metoda, jak konstrukovou validitu nástroje ověřit, a jedná se skutečně o kombinaci různých *zdrojů informací*, které by mohly nasvědčovat, že test je vzhledem ke konstrukt validní. Mezi tyto zdroje patří všechny výše popsané druhy validity a způsoby jejich odhadu, přičemž v pozadí jejich kombinace stojí teoretická úvaha o dosavadním poznání o daném

³² Statistická významnost, jak již bylo řečeno, závisí na mnoha různých faktorech, nikoliv pouze na absolutní hodnotě přírůstku R^2 .

konstruktů a jeho vztazích s dalšími konstrukty, koncepty a proměnnými (Schultz, & Whitney, 2005; Urbánek, Denglerová, & Širůček, 2011). Validizace psychodiagnostických a jiných testových metod je tedy de facto nikdy nekončícím procesem, probíhajícím souběžně s ověřováním jednotlivých typů validit konkrétního testu např. v jiných podmínkách, vůči rozdílným kritériím apod. Vzhledem k tomu, že výše uvedené kapitoly stručně shrnují základní informace potřebné i pro odhad konstruktové validity, dovolíme si tento koncept již dále nerozvádět a čtenáře se zájmem o podrobnější popis a přístupy k ověřování konstruktové validity odkázat na vybrané publikace (např. Crocker, & Algina, 2008; Urbánek, Denglerová, & Širůček, 2011).

6.5 Průnik mezi reliabilitou a validitou testu: Položková analýza

Kromě výše uvedených kritérií je možné posuzovat kvalitu konstrukce testu za pomoci položkové analýzy, která se svými interpretacemi pohybuje na pomezí konceptů reliability a validity. Z hlediska reliability je možné zkoumat homogenitu položek porovnáním jejich parametrů, z pohledu validity mohou korelace mezi položkami a celkovým skórem probandů v testu napovídat něco o jeho obsahové validitě (Schultz, & Whitney, 2005).

Při identifikaci vlastností testových položek pracuje klasická testová teorie s pojmy *obtížnost* a *citlivost* položky. *Obtížnost* položky odpovídá procentu správných odpovědí, tzn. podílu všech respondentů, kteří danou položku zodpověděli správně. Kvůli zlepšení odhadů reliability bychom dle CTT chtěli ideálně obtížnost položek pohybující se kolem 50% - při této obtížnosti je totiž možné očekávat nejvyšší variabilitu skóre. Položka by tedy neměla být ani příliš jednoduchá, ani příliš obtížná (Crocker, & Algina, 2008). Na druhou stranu můžeme namítat, že při výběru „těch nejlepších uchazečů“ bychom mohli chtít položky spíše obtížnější; ani tak by ovšem podíl respondentů, kteří položku vyřešili správně, neměl klesnout pod 20% (případně v opačném extrému přesáhnout 80%) (Urbánek, Denglerová, & Širůček, 2011). Důvod je prostý: položky, které nevyřeší téměř nikdo, stejně jako položky, které vyřeší všichni, nám nepřinesou příliš mnoho nových informací – jsou nedostatečně *citlivé*.

Citlivost říká, nakolik je položka schopná rozlišovat mezi schopnějšími a méně schopnými uchazeči, a v CTT je zjišťována v zásadě dvěma způsoby: Prostřednictvím korelace položky s celkovým skórem testu (viz níže) a pomocí rozdílu v odpovědích 27%

celkově nejúspěšnějších a 27% celkově nejméně úspěšných respondentů na položku. Ten lze vyjádřit koeficientem D , spočítaným jako:

$$D = p_u - p_l$$

kde p_u je proporce uchazečů z „nejlepších 27%“, kteří danou položku zodpověděli správně, a p_l je podíl „nejhorších 27%“ uchazečů, kteří taktéž dovedli na položku správně odpovědět. Crocker a Algina (2008) uvádějí coby přijatelný rozdíl mezi těmito skupinami $D \geq 0,4$ (případně $D \geq 0,3$ jako spodní hranici) lze se však dočíst i o koeficientech závislých na obtížnosti položky, kdy položky s obtížností 0,2-0,3 (tzn. s 20-30% celkových správných odpovědí) a 0,7-0,8 (70-80% správných odpovědí) by neměly vykazovat nižší citlivost než $D < 0,15$ a položky s obtížností mezi 30-70% $D < 0,25$ (WikiSkripta, 20. dubna 2013). Rozhodně by se pak nikde neměly objevovat koeficienty záporné, neboť takové položky „zvyhodňují“ slabší studenty (a odporují tedy účelu testu). I tehdy, je-li rozdíl v odpovědích mezi „nejlepšími“ a „nejslabšími“ respondenty příliš malý, tzn. i nejméně úspěšní uchazeči ve většině případů zodpoví položku správně nebo i nejúspěšnější respondenti nejsou schopni správnou odpověď určit, je lepší položku vyřadit; její *rozlišovací* (či také diskriminační) *schopnost* je příliš malá. To může souviset jednak s obtížností položky jako takovou (např. vyžaduje příliš specifické vědomosti), jednak s její *uhádnutelností* (např. kvůli špatně zvoleným distraktorům). Klasická testová teorie uhádnutelnost jako samostatný parametr nevyčleňuje, můžeme se s ním však setkat v teorii odpovědi na položku (Shultz, & Whitney, 2005; Urbánek, Denglerová, & Širůček, 2011). V testech SPF bychom tedy ideálně chtěli ponechat takový typ položek, jejichž obtížnost bude kolísat mezi 0,3 a 0,7 a zároveň bude jejich citlivost vyšší než 0,25 (nejlépe pak vyšší než 0,3 či 0,4).

Druhou možností odhadu citlivosti položky je její korelace s celkovým skórem uchazeče v testu. V případě dichotomických položek (které se nacházejí v testech SPF) se může jednat o modifikaci Pearsonova korelačního koeficientu, *bodově-biseriální korelaci*, počítanou pomocí vzorce na *Obr.6*.

$$r_{pb} = \frac{(\mu_1 - \mu_x)}{\sigma_x} \sqrt{p/q}$$

Obr. 6: Bodově-biseriální korelace (dle Crocker, & Algina, 2008)

Zde je μ_I je průměr celkového testového skóre osob, které danou položku zodpověděly správně, μ_x je průměrné celkové skóre všech uchazečů, σ_x je standardní odchylka celkového skóru testu, p je podíl osob, které položku zodpověděly správně a q je podíl osob, které položku zodpověděly chybně (příčemž platí že $p+q = 1$). Stejně jako u koeficientu D i u bodově-biseriální korelace chceme, aby byla její hodnota pokud možno co nejvyšší, neboť pak položka sdílí vyšší podíl variability s celkovým skórem testu – více mu „přispívá“ a lépe rozlišuje uchazeče. Problémem je, že celkový skór testu v sobě implicitně zahrnuje i skór dané položky, a tedy není možné najít naprosto nulovou korelaci, i kdyby položka testu vůbec „nepřispívala“. Proto je třeba provést korekci a korelovat položku se „zbytkem“ testu, tzn. celkovým skórem, do něhož není započítána odpověď na danou položku. Vzorec pro výpočet takto korigované bodově-biseriální korelace je na *Obr.7*, kde r_c je korigovaná korelace, r_{pb} je původní bodově-biseriální korelace, σ_x je standardní odchylka celkového skóru testu a σ_i je standardní odchylka dané položky, rovná odmocnině z $p \cdot q$ (Crocker, & Algina, 2008). Urbánek, Denglerová a Širůček (2011, 227) uvádějí, že „položka s dostatečně vysokou rozlišovací účinností by měla mít hodnotu korigované korelace (...) alespoň 0,3“.

$$r_c = \frac{r_{pb}\sigma_x - \sigma_i}{\sqrt{\sigma_i^2 + \sigma_x^2 - 2r_{pb}\sigma_x\sigma_i}}$$

Obr. 7: Korigovaná bodově-biseriální korelace (dle Crocker, & Algina, 2008)

Kromě klasické testové teorie s jejími výhodami a nevýhodami (např. závislostí na konkrétním vzorku, jeho velikosti a vlastnostech) existují i další přístupy k položkové analýze (např. teorie odpovědi na položku), jejich výpočty jsou ovšem poměrně složitější (zvláště s ohledem na fakt, aby nedošlo k jejich interpretačnímu zkreslení) a proto se v této práci omezíme pouze na položkovou analýzu prostřednictvím CTT.

Výzkumná část diplomové práce

7 Výzkumný problém, cíle práce a hypotézy

Cílem výzkumné části této práce je posouzení psychometrických vlastností testů SPF z let 2011 – 2013, konkrétně jejich reliability a validity. Za tímto účelem jsme si stanovili následující dílčí cíle a hypotézy:

1) Zjištění reliability vnitřní konzistence všech verzí testů SPF z let 2011-2013.

2) Posouzení paralelity verzí testu SPF v rámci jednotl. ročníků 2011-2013.

V rámci tohoto cíle budeme ověřovat následující hypotézy:

H1a: Pro všechny páry lichých a sudých verzí testů SPF 2011 platí, že se hodnoty jejich koeficientů vnitřní konzistence od sebe statisticky významně neliší.

H1b: Pro všechny páry lichých a sudých verzí testů SPF 2012 platí, že se hodnoty jejich koeficientů vnitřní konzistence od sebe statisticky významně neliší.

H1c: Pro všechny páry lichých a sudých verzí testů SPF 2013 platí, že se hodnoty jejich koeficientů vnitřní konzistence od sebe statisticky významně neliší.

H2a: Pro všechny páry lichých a sudých verzí testů SPF 2011 platí, že se jejich směrodatné odchylky od sebe statisticky významně neliší.

H2b: Pro všechny páry lichých a sudých verzí testů SPF 2012 platí, že se jejich směrodatné odchylky od sebe statisticky významně neliší.

H2c: Pro všechny páry lichých a sudých verzí testů SPF 2013 platí, že se jejich směrodatné odchylky od sebe statisticky významně neliší.

H3a: Pro všechny páry lichých a sudých verzí testů SPF 2011 platí, že se jejich celkové průměry od sebe statisticky významně neliší.

H3b: Pro všechny páry lichých a sudých verzí testů SPF 2012 platí, že se jejich celkové průměry od sebe statisticky významně neliší.

H3c: Pro všechny páry lichých a sudých verzí testů SPF 2013 platí, že se jejich celkové průměry od sebe statisticky významně neliší.

3) Zhodnocení obsahové validity testu SPF v rámci jednotl. ročníků 2011-2013 a posouzení obtížnosti a citlivosti různých typů položek.

4) Prověření prediktivní validity testu SPF v roce 2011 a 2012 vzhledem ke známkovému průměru studentů na konci prvního ročníku studia; srovnání inkrementální validity výsledků z testu SPF vůči dalším vybraným proměnným.

I v rámci tohoto cíle budeme ověřovat konkrétní hypotézy:

H4a: Body z testu SPF v roce 2011 statisticky významně korelují se známkovým průměrem studentů v prvním roce studia.

H4b: Body z testu SPF v roce 2012 statisticky významně korelují se známkovým průměrem studentů v prvním roce studia.

H5a: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2011) nad rámec maturitního průměru je statisticky významný.

H5b: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2012) nad rámec maturitního průměru je statisticky významný.

H6a: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2011) nad rámec zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 30% statisticky významný.

H6b: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2011) nad rámec zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 33% statisticky významný.

H6c: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2012) nad rámec zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 30% statisticky významný.

H7a: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2011) nad rámec maturitního průměru a zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 30% statisticky významný.

H7b: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2011) nad rámec maturitního průměru a zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 33% statisticky významný.

H7c: Přírůstek podílu variance známkového průměru v prvním roce studia objasněné testem SPF (2012) nad rámec maturitního průměru a zbylé části přijímací zkoušky (OT/ÚZ) je u oborů s váhou SPF 30% statisticky významný.

Kromě toho chceme zhodnotit i vzájemnou souvislost bodů z testu SPF, známkového průměru v prvním roce studia a věku.

8 Popis zvoleného metodologického rámce a metod

8.1 Typ výzkumu a metody získávání dat

V případě našeho výzkumu byl použit kvantitativní přístup a jedná se v podstatě o validizační studii. Data pro posouzení reliability (vnitřní konzistence i paralelních verzí), obsahové validity a pro položkovou analýzu testů SPF z let 2011-2013 byla získána prostřednictvím komunikace s vedoucí studijního oddělení FF UP a externí zpracovatelskou firmou, která nám na naši žádost poskytla 3 matice dat – odpovědi všech uchazečů o studium na FF UP z let 2011 až 2013, kteří absolvovali test SPF, na jednotlivé položky, spolu s uvedením číselného identifikátoru uchazeče a verze testu. Původní odpovědi (a-e) byly dále na základě vzorů správných odpovědí překódovány do podoby 0/1 (špatná či žádná/správná odpověď).

Data pro zhodnocení prediktivní validity testu SPF v roce 2011 a 2012 a porovnání s dalšími potenciálními prediktory byla získána dvěma způsoby:

Zprvce byl prostřednictvím vedoucí studijního oddělení FF UP navázán kontakt s Centrem výpočetní techniky na UP, jehož pověřený pracovník nám poskytl výpisy:

a) z databáze uchazečů o studium na FF UP za jednotlivé ročníky (2011 a 2012), které obsahovaly stejný číselný identifikátor jako matice externí zpracovatelské firmy a dále údaje, na jaký obor se uchazeč hlásil (typ, forma, obor/oborová kombinace), kolika bodů dosáhl v testu SPF a v přijímací zkoušce celkem a zda byl či nebyl přijat ke studiu, příp. z jakého důvodu;

b) z interní databáze FF UP k vahám jednotlivých částí přijímací zkoušky do Bc. studia jednotlivých oborů a jejich kombinacích v uplynulých letech (2011 a 2012);

c) z databáze současných studentů Bc. studia na FF UP se stejným číselným identifikátorem, které obsahovaly údaje k oboru či oborové kombinaci studenta, jeho studijní průměr v prvním ročníku (u dvouoborů pro oba obory dohromady) a věk.

Zadruhé byly v součinnosti s vedoucí studijního oddělení FF UP a jednotlivými studijními referentkami autorkou práce do počítače přepsány známky z maturitních vysvědčení současných studentů prvních a druhých ročníků bakalářských oborů na FF UP, pakliže tyto byly v jejich složkách k dispozici, a byl z nich pro každého studenta vypočítán maturitní průměr. Ten byl spojen se stejným číselným identifikátorem jako na výpisu z databáze současných studentů.

Veškerá výše uvedená data pak byla na základě číselného identifikátoru a údajů k jednotlivým oborům a kombinacím propojena. Vznikla výsledná matice obsahující údaje úspěšných i neúspěšných uchazečů a zapsaných studentů stran jimi absolvované přijímací zkoušky (ak. rok; obor/kombinace, typ a forma studia; složení PZk včetně vah jednotlivých částí; verze a počet bodů z testu SPF, byl-li tento absolvován; celkový počet bodů z PZk; vážený počet bodů z SPF a zbytku přijímací zkoušky), rozhodnutí o (ne)přijetí na daný obor či kombinaci a u nastoupivších studentů ještě o průměru v prvním ročníku studia, věku, popř. maturitním průměru. Tato data byla následně analyzována v souladu s cíli práce. Ukázka matice dat testů SPF je uvedena v Příloze 3.

8.2 Etické problémy a způsob jejich řešení

Ještě, než rozvedeme metody zpracování a analýzy dat, považujeme za vhodné zmínit etickou stránku našeho výzkumu. Svou povahou se totiž vymyká typickým diplomovým pracím v oblasti psychologie, především pak absencí kontaktu se „skutečnými respondenty“ – uchazeči a studenty na FF UP. V tomto ohledu jsme řešili především otázku informovaného souhlasu s přístupem k citlivým datům uchazečů/studentů a jejich zpracováním – nebylo by možné kontaktovat každého uchazeče a studenta zvlášť a ztráta dat při tomto postupu by byla prakticky neúnosná. Prostřednictvím komunikace se studijním oddělením FF UP a právním oddělením UP jsme však vyrozuměli, že při odeslání elektronické přihlášky na FF UP poskytují všichni uchazeči souhlas se statistickým zpracováním svých dat pro potřeby FF UP; stejně tak jsou zpracovávána data zapsaných studentů. Vzhledem k tomu, že jako výzkumníci jsme zároveň zaměstnanci FF UP a cíle naší práce nepřesahují anonymní statistické zpracování dat (např. uváděním citlivých osobních informací), mohla být tato klauzule vztahována i na naši práci. Zároveň jsme podepsali prohlášení o mlčenlivosti a zavázali se s daty nakládat v souladu se Zákonem č. 101/2000 Sb. o ochraně osobních údajů. Pokud jsme si vědomi, nemělo by za těchto podmínek dojít k poškození současných studentů FF UP ani bývalých uchazečů.

Zároveň jsme si vědomi citlivosti publikovaných informací vzhledem k zájmům a dobrému jménu FF UP. I z tohoto důvodu nejsou např. zveřejněny konkrétní vzorce pro výpočet vah jednotlivých částí přijímací zkoušky nebo podrobná obsahová analýza testů SPF a jednotlivých položek. Podobně si uvědomujeme, že výsledky některých analýz by potenciálně mohly sloužit jako argument nepřijatých uchazečů pro odvolací, v horším

případě pak soudní řízení. Považujeme však za nutné usilovat o zvýšení kvality přijímacího řízení, jehož součástí je i určitá sebereflexe co do plnění účelu PZk a míra transparency tohoto procesu vůči uchazečům. Vycházíme z toho, že i jiné velké české a zahraniční univerzity publikují validizační studie svých testů studijních předpokladů, a řídíme se jejich příkladem.

8.3 Metody zpracování a analýzy dat

K naplnění našich cílů byly obecně použity metody deskriptivní a inferenční statistiky; pro větší přehlednost uvádíme jejich konkrétní podoby v níže rozvedených podkapitolách. Vzhledem k povaze proměnných a velikosti datového souboru/vzorku byly s odvoláním na centrální limitní větu použity parametrické testy a není-li uvedeno jinak, byly hypotézy testovány oboustranně na hladině $\alpha = 0,05$. Data byla zpracována za pomoci programů Microsoft Excel 2003, Remark Office OMR 8 a SPSS for Windows, v.17.

8.3.1 Posouzení reliability testů SPF 2011-2013

Pro posouzení reliability testů SPF z let 2011-2013 byly nejprve pro jednotlivé verze vypočítány průměrné počty správně zodpovězených otázek (vč. minima a maxima), jejich směrodatné odchylky a koeficienty vnitřní konzistence pomocí Kuder-Richardsonova vzorce 20 (KR-20). Srovnání párů korelačních koeficientů (H1a-c) se opíralo o z-test, přičemž původní korelační koeficienty bylo třeba nejprve transformovat pomocí Fischerovy Z-transformace dle rovnice na *Obrázku 8*, kde Z je transformovaná hodnota korelačního koeficientu, r je původní korelační koeficient a \ln je přirozený logaritmus.

$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

Obr. 8: Fischerova Z-transformace (dle Eid, Gollwitzer, & Schmitt, 2010)

Takto upravené korelace bylo možné porovnat na základě vzorce uvedeném na *Obrázku 9*, kde Z_1 a Z_2 jsou transformované hodnoty jednotlivých korelací a $\sigma_{(Z_1-Z_2)}$ označuje rozptyl vzájemného rozložení těchto hodnot, jehož výpočet je uveden na *Obrázku 10* s n coby počtem hodnot/uchazečů absolvujících danou verzi.

$$Z = \frac{Z_1 - Z_2}{\sigma_{(Z_1-Z_2)}}$$

Obr. 9: Vzorec pro z-test k porovnání dvou korelačních koeficientů (tamtéž)

$$\sigma_{(z_1-z_2)} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}$$

Obr. 10: Rozptyl rozložení dvou korelačních koeficientů (tamtéž)

Vzhledem k tomu, že v jednotlivých ročnících testu SPF bylo provedeno více porovnávání (konkrétně 8 srovnání pro 16 verzí v letech 2011 jakož i 2012 a 6 srovnání pro 12 verzí v roce 2013), vystavili jsme se tím vyšší šanci najít falešně pozitivní výsledek, tzv. *family-wise error* (Eid, Gollwitzer, & Schmitt, 2010). Její odhad je znázorněn pomocí rovnice

$$\alpha_{fam} = 1 - (1 - \alpha)^s$$

kde α_{fam} odpovídá kumulované chybě alfa, α hladině alfa zvolené pro každé porovnání a s počtu provedených srovnání. Z tohoto důvodu byla před jednotlivými porovnáními potřeba korekce hladiny α_r vzhledem k plánovaným srovnáním, k čemuž bylo využito metody podle Šidáka (1967, in Eid, Gollwitzer, & Schmitt, 2010):

$$\alpha_r = 1 - \sqrt[s]{1 - \alpha}$$

Výsledky těchto analýz jsou shrnuty v Tabulkách 4 a 5 v části Výsledky. Srovnání průměrů párových verzí (H2a-c) bylo provedeno dvouvýběrovým t-testem, srovnání směrodatných odchylek (H3a-c) proběhlo pomocí F-testu pro rozptyl, přičemž byla aplikována stejná pravidla ke korekci kumulace alfa-chyby jako v případě porovnávání korelačních koeficientů.

Jsme si vědomi, že tento postup není standardním ověřením reliability paralelních verzí, které běžně probíhá na datech týchž uchazečů, kteří prošli oběma/všemi verzemi. V podmínkách administrace testů SPF však není možné taková data získat. Přesto existují obsahově totožné – „párové“ – verze (viz kap. 6.3.1) testů administrované ve stejné skupině uchazečů, které by teoreticky měly vykazovat velmi podobné parametry (průměry, směrodatné odchylky, popř. vnitřní konzistenci)³³. Proto jsme se rozhodli alespoň pro počáteční „hrubý odhad“ za pomoci výše uvedených metod.

³³ „Liché“ verze se naproti tomu mohou svými parametry navzájem do určité míry lišit, neboť skupiny uchazečů, jež je absolvovaly, byly pokaždé „jinak složené“ – viz kap. 6.1 a 6.3.1. I proto by srovnání jejich průměrů a rozptylů pomocí např. jednofaktorové analýzy variance odráželo i potenciálně odlišné pravé skóry uchazečů, které by nebylo možné odlišit od chyby měření.

8.3.2 Obsahová validita a položková analýza

Jak již bylo diskutováno v teoretické části práce (kap. 6.4.1), je zhodnocení obsahové validity testů SPF problematické. Jelikož posouzení panelem expertů je de facto součástí již samotného procesu tvorby testů, rozhodli jsme se v této práci omezit na statistické zdroje důkazů o obsahové validitě, opřené o faktorovou a položkovou analýzu.

V souladu s doporučeními v odborné literatuře (Eid, Gollwitzer, & Schmitt, 2010; Schultz, & Whitney, 2005) byla ke zjištění dimenzionality testů zvolena metoda explorační faktorové analýzy (EFA), neboť si na základě teoretických východisek nejsme počtem dimenzí jisti (a tedy nepoužíváme konfirmační FA). Zároveň byla metoda EFA upřednostněna před analýzou hlavních komponent (angl. *principal component analysis*, PCA), neboť na rozdíl od PCA orientované na celkovou variabilitu dat slouží EFA k odhalení latentních proměnných a je tedy pro případ analýzy testů typu SPF vhodnější (Schultz, & Whitney, 2005). Předpokladem EFA je dostatečně velký vzorek – jak obecně ($n > 100$), tak relativně vůči počtu položek ($n:k = 5:1$) (Schultz, & Whitney, 2005). V našem případě by tedy pro každou verzi mělo ideálně platit $n > 250$, minimálně pak $n > 100$.

Byl také zadán Bartlettův test sféricity a zobrazení Kaiser-Meyer-Olkinova koeficientu (KMO) k posouzení míry vhodnosti faktorové analýzy. Signifikance Bartlettova testu značí, že mezi položkami jsou dostatečně velké interkorelace, aby provedení FA mělo smysl, test je však náchylný na počet pozorování ve vzorku (při vyšším n jsou i relativně nízké korelace významné). Proto Škaloudová (2010) doporučuje posouzení podle KMO, kdy $KMO > 0,8$ značí vysokou vhodnost FA, $KMO = 0,6-0,8$ nízkou až střední vhodnost FA a při $KMO < 0,5$ nemá dle autorky provedení faktorové analýzy smysl.

Jako metoda extrakce faktorů byla - opět v souladu se zaběhnutou praxí ve společenských vědách – zvolena analýza hlavních os (angl. *principal axis factoring*) na základě matice korelací mezi položkami. Maximální počet iterací byl u všech verzí a ročníků nastaven na 50 (odpovídající celkovému počtu položek). Počet faktorů byl vybírán jednak pomocí Kaiserova kritéria pro vlastní hodnoty (angl. *eigenvalues*) > 1 , jednak po posouzení sutinového grafu, neboť Kaiserovo kritérium má tendenci nadhodnocovat počet faktorů (Eid, Gollwitzer, & Schmitt, 2010). Urbánek, Denglerová a Širůček (2011) uvádějí, že položky v unidimenzionálním testu by měly v (prvním) společném faktoru vykazovat náboj alespoň 0,3.

Zároveň jsme po prvotní extrakci faktorů použili neortogonální rotaci typu Promax, neboť jsme předpokládali, že faktory by mezi sebou mohly být korelované a jejich rotování by mohlo usnadnit jejich interpretaci. Rotace Promax se v prvním kroku zakládá na ortogonální rotaci Varimax, která zvýší variabilitu uvnitř jednotlivých sloupců, takže každý faktor koreluje silně jen s určitou skupinou proměnných. Ve druhém kroku pak vznikne matice rotovaných faktorů, v níž se odrazí i případná interkorelace mezi faktory. Faktor Kappa pro rotaci Promax byl v souladu s doporučením v literatuře nastaven na $\kappa = 4$ (Schultz, & Whitney, 2005).

Položková analýza jednotlivých verzí testů SPF 2011 až 2013 byla provedena dle postupů klasické testové teorie (CTT). Byly vypočteny korigované bodově biseriální korelace jednotlivých položek s celkovým skórem testu, procentuální poměry, kolikrát byly zvoleny jednotlivé distraktory a správná odpověď, celková obtížnost položky (jako procentuální podíl všech uchazečů, kteří danou položku zodpověděli správně) a podíl z neúspěšnějších 27% a nejméně úspěšných 27% uchazečů, kteří na danou položku odpověděli správně. Z nich byl dále vyjádřen koeficient D (rozdíl mezi proporcemi 27% neúspěšnějších a 27% nejméně úspěšných uchazečů) pro odhad citlivosti položky, který by u velmi obtížných (20-30% celkových správných odpovědí) a velmi snadných (70-80% správných odpovědí) položek vykazovat nižší hodnotu než $D < 0,15$ u položek s obtížností mezi 30-70% pak $D < 0,25$ a ideálně by měl nabývat vyšších hodnot než $D > 0,3$ (WikiSkripta, 20. dubna 2013). Stejně tak bodově biseriální korelace by měla nabývat hodnot alespoň $r = 0,3$ (Urbánek, Denglerová, & Širůček, 2011). Tyto parametry byly využity k identifikaci typově potenciálně vhodných a nevhodných položek k zařazení (v obměněné podobě) v dalších ročnících SPF.

Uvědomujeme si, že parametry položek vypočtené na základě CTT mohou být potenciálně zkreslené konkrétním vzorkem uchazečů, na němž byly odhadnuty. Vzhledem k nejasné dimenzionalitě testů a možným interpretačním zkreslením při použití sofistikovanějších výchozích teorií (např. teorie odpovědi na položku) jsme se však pro prvotní orientaci ve vlastnostech položek přiklonili právě k tomuto postupu.

8.3.3 Kritériová prediktivní a inkrementální validita

Z hlediska kritériové validity nás zajímala především prediktivní schopnost testu SPF vůči známkovému průměru studentů v prvním roce studia (FYGPA), který je v odborné literatuře na podobné téma častým kritériem akademického úspěchu (např. Camara,

2004a). Jelikož pro studenty, kteří test SPF absolvovali v roce 2013, ještě není průměr známek za první ročník znám, omezily se naše výpočty pouze na data uchazečů a studentů z let 2011 a 2012. Kde to bylo možné, opírali jsme se o největší dostupný výběrový soubor (ne všechny údaje totiž byly dostupné od všech uchazečů a studentů) a využili korekci pro omezení variability.

Za účelem zjištění obecné souvislosti mezi body v testu SPF a FYGPA (H4a-b) byla vypočítána Pearsonova korelace mezi body z testu SPF u všech přijatých uchazečů, u nichž byly tyto body známy (tzn. byli přijeti na základě PZk i na odvolání), bez ohledu na typ a formu studia a obor, na který se hlásili, i bez ohledu na absolvovanou verzi testu. Naším východiskem k tomuto postupu bylo, že při předpokládané ekvivalenci verzí a obsahové obhajitelnosti konstruktů studijních předpokladů i jeho naplnění v testech SPF by skóre v testu SPF měl předikovat známky v prvním ročníku (FYGPA) napříč všemi obory i verzemi konzistentně. Korelace mezi body testu SPF a FYGPA tedy byla vypočtena na nejvyšším možném dosažitelném vzorku (viz také kap. 8.4 Výzkumný soubor) a dále byla provedena korekce omezení variability dle vzorce na Obr. 5 v kapitole 6.4.2. Směrodatná odchylka neomezeného souboru byla vypočtena jako směrodatná odchylka bodů získaných všemi účastníky testu SPF v daném ročníku, směrodatná odchylka omezeného souboru se poté vztahovala k počtu studentů přijatých v daném roce na základě PZk a na odvolání.

Pro srovnání byly vypočteny ještě Pearsonovy korelace mezi maturitním průměrem všech zapsaných uchazečů, u nichž byl tento znám, a FYGPA (pro studenty nastoupivší v roce 2011 a 2012) a mezi věkem studentů a jejich známkovým průměrem v prvním ročníku. Stejně tak byl pomocí jednofaktorové analýzy variance srovnán FYGPA uchazečů přijatých na základě přijímací zkoušky, na odvolání a bez přijímací zkoušky. Z podobné logiky jako v předchozím případě jsme i zde zanedbali typ, formu a obor studia, i když uznáváme, že oddělené analýzy pro jednotlivé obory či kombinace³⁴ by mohly být přesnější. Vzhledem k zaměření této práce a co nejmenšímu omezení výběrového souboru (u mat. průměru a věku nejsou k dispozici „populační“ směrodatné odchylky pro korekci omezení variability) jsme se však spokojili s tímto hrubým odhadem.

Zároveň nás zajímal přínos testu SPF nad rámec ostatních, v literatuře často zmiňovaných prediktorů, a to oborového testu a maturitního průměru uchazeče. Za tímto

³⁴ Ideálem by pak byla dostupnost dat-známeček za jednotlivé předměty každého studenta, jak to doporučují Sackett, Borneman a Connelly (2008).

účelem byla – v každém z relevantních ročníků (2011 a 2012) zvlášť - provedena sada krokových mnohonásobných lineárních regresí se známkovým průměrem v prvním ročníku jako závislou proměnnou a byla ověřována signifikance přírůstku R^2 čili *koeficientu determinace* po zadání/zohlednění příslušných prediktorů:

- a) Pro ověření hypotéz H5a-b byl v každém ročníku PZk sestaven soubor studentů, kteří absolvovali test SPF (bez ohledu na verzi a jeho váhu v PZk) a zároveň byl znám jejich maturitní průměr. V prvním kroku regrese byl jako prediktor FYGPA zadán maturitní průměr, ve druhém kroku pak skóre v testu SPF. K doplnění informací byl ve třetím kroku zadán ještě věk studenta, neboť jeho významnost v početně omezeném vzorku (oproti výše uvedené jednoduché Pearsonově korelaci) se mohla lišit.
- b) Pro ověření hypotéz H6a-c a H7a-c byl v každém ročníku sestaven soubor studentů, kteří u PZk kromě testu SPF absolvovali ještě oborový test, ústní pohovor a/nebo dokládali další (např. jazykové) certifikáty a osvědčení, a pro něž byl zároveň znám jejich maturitní průměr. Do tohoto souboru byli zahrnuti pouze ti studenti, jejichž maximální dosažitelný počet bodů u přijímací zkoušky a váhy jednotlivých částí PZk ve výpisu z databáze studentů odpovídaly specifikacím jejich oboru ve výpisu z interní databáze FF UP (viz kap. 8.4) Jednalo se tedy o studenty, kteří u přijímací zkoušky mohli získat maximálně 100 bodů a poměry vah testu SPF a zbylé části zkoušky činily 30:70. Jelikož u početně významné skupiny uchazečů/studentů psychologie v roce 2011 byly váhy stanoveny na 33:67, byla tato skupina vyčleněna samostatně a podrobena regresním analýzám zvlášť. Pro všechny testované soubory však byly vytvořeny stejné modely regresní analýzy, do nichž pouze vstupovaly různě vážené skóre z testu SPF a zbylé části zkoušky. Do prvního modelu (ověření H6) byl v prvním kroku zadán (vážený) výsledek „zbylé části PZk“, ve druhém kroku vážený skóre testu SPF a pro doplnění informací ještě ve třetím kroku maturitní průměr. Do druhého modelu (ověření H7) byl nejprve zadán maturitní průměr, poté ve druhém kroku výsledek „zbylé části PZk“ (s příslušnou váhou), ve třetím kroku vážený skóre SPF a ve čtvrtém kroku pro úplnost ještě věk studenta.

Srovnáním přírůstku R^2 v jednotlivých modelech a porovnáním standardizovaných beta-koeficientů jednotlivých prediktorů pak mohl být zhodnocen přínos každé proměnné ke (kvalitě) předpovědi známkového průměru v prvním roce studia.

8.4 Výzkumný soubor

V teoretické části práce jsme se zabývali otázkou, koho lze vlastně považovat za cílovou populaci testů SPF, zda všechny uchazeče o studium na FF UP či všechny „potenciální uchazeče“, tzn. osoby přiměřeného věku, které by se na FF UP hlásit mohly. Je ovšem zásadní si uvědomit, že cílem této práce nejsou inference o schopnostech uchazečů, nýbrž o vlastnostech *testů* SPF. Základním souborem by se tedy daly označit veškeré testy SPF, které byly na FF UP doposud vytvořeny a potenciálně i veškeré testy vytvořené v budoucnu, pakliže budou mít podobnou strukturu, způsob administrace, vyhodnocení a interpretace jako doposud. Pro popis testů SPF z let 2011-2013 odkazujeme na kapitolu 5.

Soubor dat, s nimiž jsme pracovali, lze de facto považovat za totální výběr ze všech testů SPF, které doposud existují, resp. proběhly (v letech 2011-2013). Pro účely ověření reliability, položkové a faktorové analýzy a korekce bodů SPF pro omezení variability jsme využili data 4.480 uchazečů v souboru SPF 2011, 4.459 uchazečů v souboru SPF 2012 a 4.231 uchazečů v souboru SPF 2013. Podrobné údaje k jednotlivým verzím lze nalézt v kapitole Výsledky v Tab. 4.

Pro výpočty prediktivní validity byly vytvořeny soubory dat na základě kritérií popsanych v kap. 8.3.3. Zahrnuta tedy byla v první řadě pouze data účastníků (a zapsaných studentů) SPF 2011 a SPF 2012, neboť pro uchazeče z akad. roku 2013/2014 nejsou v současnosti ještě dostupné známkové průměry za první ročník studia. V rámci těchto souborů byli do výpočtů zahrnuti pouze přijatí a zapsaní studenti, pro něž byl k dispozici známkový průměr alespoň v prvním roce studia a alespoň tři z následujících proměnných (neboť proměnná „věk“ a „rozhodnutí o přijetí“ byla známa pro všechny zapsané uchazeče):

- a) rozhodnutí o přijetí (PZk, bez PZk, na odvolání);
- b) body z testu SPF;
- c) body z oborového testu a/nebo ústního pohovoru a další části PZk, pakliže byly tyto části přijímací zkoušky vyžadovány;
- d) maturitní průměr;
- e) věk.

Tab. 3 shrnuje počty uchazečů, kteří uvedené podmínky splňují, stejně jako jejich počty v jednotlivých souborech dělených dle testovaných proměnných. U počtů zapsaných

uchazečů, kteří zároveň absolvovali ještě jiné části přijímací zkoušky (OT/ÚZ atp.) než test SPF, jsou v příslušných kolonkách dva údaje: První údaj (bez závorek) odkazuje na počet uchazečů, s nimiž bude v příslušných analýzách počítáno, neboť byla dodržena kritéria o bodovém poměru SPF:OT/ÚZ, která byla popsána ve Zprávě o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2011/2012 (nedat.) a Zprávě o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2012/2013 (nedat.) a tato kritéria se zároveň shodovala s výpisem z interní databáze FF UP. Jednalo se o váhy 30% u testu SPF a 70% u zbylé části přijímací zkoušky s maximálním součtem 100 bodů celkem, s výjimkou uchazečů o studium psychologie (prezenční i kombinované) v roce 2011. Zde se jednalo o poměr 33% u testu SPF a 67% u zbytku (dvoukolové) zkoušky s max. počtem 100 bodů; velikost i povaha této skupiny však pro nás byla z výzkumného hlediska natolik zajímavá, že jsme se rozhodli pro ni vyčlenit samostatné regresní analýzy (jinak obsahově shodné s těmi pro soubor „30:70“). Důvodem, proč bylo třeba omezit počet analyzovaných účastníků/studentů, byly některé nesrovnalosti, která vykazovala data ve výpisech z jednotlivých databází CVT (např. maximální dosažitelný počet více než 100 bodů, jiné faktické váhy jednotlivých částí PZk při započtení různých certifikátů apod.), jež by nebylo možné statisticky ošetřit a které by způsobovaly zkreslení v regresních rovnicích. Celkové počty uchazečů, kterých by se analýzy s OT/ÚZ týkaly, pakliže by data tyto nesrovnalosti neobsahovala, jsou v Tab. 3 uvedeny kurzívou v závorce. Seznam oborů s „nesrovnalostmi“ je k nalezení v Příloze 4. Počty zapsaných uchazečů, u nichž byly známy výsledky testů SPF (bez ohledu na jejich váhu v celkovém hodnocení, viz výše), v sobě zahrnují i počty uchazečů přijatých na odvolání, neboť i u nich chceme znát, jak dobře body SPF dovedou předikovat známkový průměr v prvním roce VŠ studia. Ze stejného důvodu zahrnují počty uchazečů se známým maturitním průměrem i uchazeče přijaté bez přijímací zkoušky.

Tab. 3: Data uchazečů v jednotlivých souborech SPF splňujících podmínky pro výpočty prediktivní validity

	SPF 2011	SPF 2012
Přijatých uchazečů (celkem)	1.250	1.331
z toho:		
na základě přijímací zkoušky	1.207	1.310
bez přijímací zkoušky	27	13
na základě odvolání	16	8
<i>do prezenčního studia</i>	1.048	1.140
<i>na jednoobory</i>	708	751
<i>na dvouobory</i>	340	389
<i>do kombinovaného studia</i>	202	191
<i>na jednoobory</i>	178	172
<i>na dvouobory</i>	24	19
Dle dostupných proměnných:		
body SPF	1.223	1.318
maturitní průměr	821	973
věk	1250	1331
body SPF + maturitní průměr + věk	797	959
body SPF + OT/ÚZ + maturitní průměr + věk	435	433
z toho:	(578)	(602)
SPF s váhou 30%	345	433
SPF s váhou 33%	90	X

Pozn.: Údaje psané kurzívou slouží pouze pro orientační přehled a v analýzách s nimi nebude počítáno.

9 Výsledky

9.1 Výsledky posouzení reliability testů SPF 2011-2013

Základní deskriptivní statistiky stejně jako koeficienty vnitřní konzistence jednotlivých verzí testu v letech 2011 až 2013 jsou shrnuty v Tab. 4 na následující stránce. Verze testů SPF 2011 dosahují koeficientů vnitřní konzistence v rozmezí hodnot 0,67-0,78; to lze vzhledem ke kritériu spodní hranice uspokojivé testové reliability dle Schultze a Whitneyho (2005) $r = 0,7$ považovat za uspokojivé. V roce 2012 dosahují verze testů SPF koeficientů vnitřní konzistence v rozmezí hodnot 0,60-0,76; lze mezi nimi tedy pozorovat relativně vyšší rozptyl než v předchozím ročníku, přičemž jednotlivé verze se zdají být méně homogenní. Konečně v roce 2013 nabývají koeficienty vnitřní konzistence jednotlivých verzí testu SPF hodnot v rozmezí 0,60-0,73; tento trend se podobá hodnotám z roku 2012 a lze jej považovat spíše za slabší vzhledem ke kritériu uváděnému Schultzem a Whitneym (2005).

Tabulka 5 na straně 77 shrnuje údaje potřebné pro zhodnocení hypotéz H1a-c. Velikost family-wise error při hladině $\alpha = 0,05$ činí pro roky 2011 a 2012 (8 párových porovnání) $\alpha_{fam} = 1 - (1 - 0,05)^8 = 0,34$ a pro rok 2013 (6 porovnání) $\alpha_{fam} = 1 - (1 - 0,05)^6 = 0,27$. Proto by hypotézy H1a a H1b měly být hodnoceny na korigované hladině $\alpha_r = 1 - \sqrt[8]{(1 - 0,05)} = 0,006$ a hypotéza H1c na hladině $\alpha_r = 1 - \sqrt[6]{(1 - 0,05)} = 0,008$. Jak je ovšem z Tabulky 5 patrné, hodnoty koeficientu z pro všechny páry verzí ve všech ročnících nejsou významné ani na hladině $\alpha = 0,05$. Proto **přijímáme hypotézy H1a, H1b a H1c**, totiž že pro všechny páry lichých a sudých verzí testů SPF 2011, 2012 a 2013 platí, že se hodnoty jejich koeficientů vnitřní konzistence od sebe statisticky významně neliší.

Výsledky dvouvýběrového t-testu pro srovnání průměrů párových verzí SPF v letech 2011, 2012 a 2013 (H2a-c) jsou stejně jako výsledky F-testu pro rozptyl (H3a-c) uvedeny v Tabulce 6 na straně 78. Hypotézy H2a a H2b byly pro mnohonásobná porovnávání opět testovány na hladině významnosti $\alpha_r = 0,006$. Na této hladině nejsou rozdíly v průměrech párových verzí testů SPF 2011 ani 2012 signifikantní, ačkoliv v SPF 2011 by na hladině $\alpha = 0,05$ byly významné rozdíly mezi verzemi 9 a 10 a verzemi 11 a 12, v SPF 2012 by pak byl významný rozdíl mezi verzemi 9 a 10 a verzemi. Přesto **přijímáme hypotézy H2a a H2b**. Hypotéza H2c byla testována na hladině významnosti $\alpha_r = 0,008$. Ani v tomto případě nebyl na této hladině shledán mezi průměry párových verzí rozdíl, ačkoliv na hladině $\alpha = 0,05$ by byl sig. rozdíl mezi verzemi 7 a 8. Přesto **přijímáme hypotézu H2c**.

Tab. 4: Základní statistiky jednotlivých verzí testů SPF v letech 2011-2013

Verze testu	2011					2012					2013					
	Počet test. os.	Min - Max	Průměr (SD)	KR-20	Počet test. os.	Min - Max	Průměr (SD)	KR-20	Počet test. os.	Min - Max	Průměr (SD)	KR-20	Počet test. os.	Min - Max	Průměr (SD)	KR-20
1	584	13 - 45	28,44 (5,74)	0,71	431	11 - 40	27,18 (5,38)	0,69	367	9 - 42	27,01 (5,18)	0,65			27,01 (5,18)	0,65
2	558	12 - 47	28,06 (6,31)	0,76	422	3 - 42	26,64 (5,11)	0,65	374	11 - 42	26,55 (5,22)	0,65			26,55 (5,22)	0,65
3	333	10 - 46	30,79 (6,13)	0,76	308	13 - 43	28,71 (4,66)	0,60	375	10 - 38	25,98 (4,96)	0,60			25,98 (4,96)	0,60
4	336	15 - 46	30,23 (6,04)	0,75	319	15 - 42	28,75 (4,81)	0,64	362	10 - 43	26,26 (5,19)	0,63			26,26 (5,19)	0,63
5	257	13 - 41	27,0 (5,90)	0,73	248	10 - 42	27,23 (5,32)	0,68	125	13 - 42	28,15 (5,61)	0,73			28,15 (5,61)	0,73
6	245	7 - 41	26,62 (5,97)	0,74	230	10 - 40	27,77 (6,03)	0,76	130	15 - 41	27,89 (5,60)	0,73			27,89 (5,60)	0,73
7	100	13 - 38	24,24 (5,74)	0,72	252	13 - 42	27,92 (5,88)	0,73	344	9 - 39	24,76 (5,55)	0,68			24,76 (5,55)	0,68
8	93	12 - 36	23,77 (5,38)	0,68	275	14 - 45	28,28 (5,82)	0,73	624	9 - 43	25,60 (5,65)	0,70			25,60 (5,65)	0,70
9	357	15 - 46	31,23 (5,87)	0,77	194	9 - 40	27,75 (5,58)	0,71	342	15 - 40	26,73 (5,07)	0,64			26,73 (5,07)	0,64
10	335	14 - 45	30,34 (5,92)	0,78	166	11 - 41	26,21 (5,84)	0,74	644	8 - 42	26,56 (5,48)	0,70			26,56 (5,48)	0,70
11	360	14 - 41	27,76 (5,55)	0,68	307	8 - 39	24,97 (5,71)	0,71	276	13 - 45	29,12 (5,75)	0,72			29,12 (5,75)	0,72
12	340	11 - 43	26,70 (5,86)	0,72	351	6 - 44	24,72 (5,73)	0,72	268	16 - 44	29,23 (5,54)	0,69			29,23 (5,54)	0,69
13	345	14 - 45	28,4 (5,60)	0,72	213	13 - 41	27,94 (4,81)	0,61	X	X	X	X			X	X
14	341	14 - 41	28,68 (5,15)	0,67	196	8 - 39	26,79 (4,89)	0,62	X	X	X	X			X	X
15	151	14 - 40	27,71 (5,48)	0,72	272	12 - 43	25,61 (5,24)	0,67	X	X	X	X			X	X
16	145	12 - 42	26,97 (5,15)	0,68	275	13 - 40	25,37 (5,28)	0,68	X	X	X	X			X	X
Celkem	4.480	7 - 47	28,45 (6,05)	X	4.459	3 - 45	26,94 (5,53)	X	4.231	8 - 45	26,67 (5,52)	X			26,67 (5,52)	X

Pozn.: Min-Max, průměr i směrodatná odchylka (SD) se vztahují k počtu správně zodpovězených otázek v testu.

Tab. 5: Srovnání koeficientů vnitřní konzistence párových verzí testu SPFF 2011-2013

Pár verzí	2011						2012						2013					
	Z1	Z2	SD	z	p		Z1	Z2	SD	z	p		Z1	Z2	SD	z	p	
1-2	0,89	1,00	0,06	-1,84	0,07		0,85	0,78	0,07	1,06	0,29		0,78	0,78	0,07	0,00	1,00	
3-4	1,00	0,97	0,08	0,30	0,77		0,69	0,76	0,08	-0,81	0,42		0,69	0,74	0,07	-0,66	0,51	
5-6	0,93	0,95	0,09	-0,24	0,81		0,83	1,00	0,09	-1,81	0,07		0,93	0,93	0,13	0,00	1,00	
7-8	0,91	0,83	0,15	0,54	0,59		0,93	0,93	0,09	0,00	1,00		0,83	0,87	0,07	-0,57	0,57	
9-10	1,02	1,05	0,08	-0,33	0,74		0,89	0,95	0,11	-0,59	0,55		0,76	0,87	0,07	-1,62	0,11	
11-12	0,83	0,91	0,08	-1,03	0,30		0,89	0,91	0,08	-0,26	0,80		0,91	0,85	0,09	0,69	0,49	
13-14	0,91	0,81	0,08	1,26	0,21		0,71	0,73	0,10	-0,16	0,87		X	X	X	X	X	
15-16	0,91	0,83	0,12	0,67	0,50		0,81	0,83	0,09	-0,21	0,83		X	X	X	X	X	

Tab. 6: Výsledková tabulka pro t-test a F-test párových verzí SPF 2011-2013

Pár verzí	Pr.1	Pr.2	t	sv	p	N1	N2	SD1	SD2	F	p
SPF 2011											
1-2	28,44	28,06	1,063	1140	0,28	584	558	5,74	6,31	1,208	0,024
3-4	30,79	30,23	1,190	667	0,23	333	336	6,14	6,05	1,029	0,79
5-6	27,00	26,62	0,708	500	0,48	257	245	5,91	5,98	1,023	0,86
7-8	24,24	23,77	0,578	191	0,56	100	93	5,77	5,41	1,137	0,53
9-10	31,23	30,33	1,988	690	0,047	357	335	5,88	5,93	1,019	0,86
11-12	27,76	26,70	2,465	698	0,014	360	340	5,56	5,86	1,114	0,32
13-14	28,40	28,68	-0,688	684	0,49	345	341	5,61	5,16	1,181	0,12
15-16	27,71	26,97	1,198	294	0,23	151	145	5,49	5,17	1,130	0,46
SPF 2012											
1-2	27,18	26,65	1,478	851	0,14	431	422	5,38	5,12	1,106	0,30
3-4	28,71	28,75	-0,101	625	0,92	308	319	4,67	4,82	1,066	0,58
5-6	27,23	27,77	-1,037	476	0,30	248	230	5,33	6,04	1,285	0,054
7-8	27,92	28,28	-0,711	525	0,48	252	275	5,89	5,83	1,022	0,86
9-10	27,75	26,21	2,552	358	0,011	194	166	5,59	5,86	1,097	0,54
11-12	24,97	24,72	0,552	656	0,58	307	351	5,72	5,74	1,008	0,95
13-14	27,94	26,79	2,397	407	0,017	213	196	4,83	4,90	1,030	0,83
15-16	25,61	25,37	0,523	545	0,60	272	275	5,25	5,29	1,016	0,90
SPF 2013											
1-2	27,01	26,55	1,18	739	0,24	367	374	5,19	5,23	1,014	0,89
3-4	25,98	26,26	-0,750	735	0,45	375	362	4,97	5,19	1,093	0,39
5-6	28,15	27,89	0,368	253	0,71	125	130	5,63	5,62	1,005	0,98
7-8	24,77	25,60	-2,202	966	0,028	344	624	5,56	5,66	1,035	0,73
9-10	26,73	26,60	0,481	984	0,63	342	644	5,08	5,49	1,166	0,11
11-12	29,12	29,23	-0,238	542	0,81	276	268	5,76	5,56	1,076	0,55

Pozn.: Pr.1 a Pr.2 = průměrný počet správně zodpovězených otázek v příslušných verzích; sv = stupně volnosti; N1 a N2 = počet účastníků jednotliv. verzí; SD = směrodatné odchylky. Kurzívou zvýrazněné p-hodnoty jsou signifikantní na hladině alfa = 0,05.

Hypotézy H3a a H3b o rovnosti rozptylů párových verzí SPF v letech 2011 a 2012 byly opět testovány na hladině významnosti $\alpha_r = 0,006$. Jak je patrné z Tab. 6, nejsou na této hladině významnosti mezi jednotlivými párovými verzemi v obou ročnících

signifikantní rozdíly, ačkoliv rozdíl v rozptylu verzí 1 a 2 v SPF 2011 by byl statisticky významný na hladině $\alpha = 0,05$. Přesto **přijímáme hypotézy H3a a H3b**. Rovnost rozptylů párových verzí SPF 2013 byla ověřována na hladině významnosti $\alpha_r = 0,008$. Ani v tomto případě nebyl na této hladině shledán mezi rozptyly párových verzí rozdíl, a tedy **přijímáme hypotézu H3c**.

Celkově lze shrnout, že na uvedených hladinách alfa nebyl shledán rozdíl mezi párovými verzemi testů SPF v jednotlivých ročnících, a to jak v jejich vnitřních konzistencích, tak ani v jejich průměrech a rozptylech.

9.2 Výsledky zhodnocení obsahové validity a položkové analýzy SPF

Faktorová analýza ke zhodnocení obsahové validity testů SPF z let 2011 – 2013 proběhla v několika krocích. Nejprve byly ověřeny předpoklady pro provedení explorační faktorové analýzy (EFA), totiž posouzení počtu pozorování v jednotlivých verzích (min. $n > 100$, ideálně pak $n > 250$) a zhodnocení Kaiser-Meyer-Olkinova (KMO) koeficientu; tyto údaje shrnuje Tabulka III v Příloze 5. Hodnoty KMO se napříč ročníky vesměs pohybovaly na spodní hranici střední až nízké vhodnosti FA podle Škaloudové (2010). Pod tuto mez klesaly především u verzí, kde počet účastníků nedosahoval požadovaného $n > 250$; v těchto případech tedy v datech nebyla objevena uspokojivě vysoká vzájemná korelace mezi položkami. Pro úplnost jsme však s EFA pokračovali i v těchto verzích, s vědomím omezení interpretačních možností výsledků.

V dalším kroku jsme pomocí metody hlavních os provedli prvotní odhad počtu latentních faktorů. Na podkladě Kaiserova kritéria pro vlastní hodnoty > 1 bylo ve všech ročnících SPF z každé verze vyextrahováno 19 – 22 faktorů (viz prototypická Tab. IV v Příloze 5). V Příloze 5 je dále uveden příklad hodnot komunalit jednotlivých položek – čili jejich variance vysvětlené vyextrahovanými (cca dvaceti) faktory. Z Tab. IV a V (tamtéž) je patrné, že podíly variance jednotlivých položek vysvětlené nalezenými faktory nejsou nikterak vysoké a zřídka kdy přesahovaly kumulativní hodnoty (i při nalezených 22 faktorech s vlastním číslem > 1) 40%. Zdá se tedy, že v datech je vysoký podíl residuální variance, který není ani prostřednictvím těchto faktorů zachycen. Z pohledu na sutinové grafy (viz Graf I, tamtéž) navíc bylo zřejmé, že pro analýzu jsou v každé verzi vhodné maximálně dva, při velmi liberálním přístupu tři faktory.

Pro další krok – rotaci metodou Promax – jsme tedy zvolili pouze dva faktory. Podíly variance v jednotlivých verzích, které jsou těmito dvěma faktory objasněny (před a po rotaci) shrnuje Tab. 7. Zde také nalezneme údaje o vzájemné korelaci mezi oběma faktory.

Tab. 7: Interkorelace faktorů 1 a 2 a jimi vysvětlená variance v testech SPF 2011-2013

Verze testu	SPF 2011			SPF 2012			SPF 2013		
	r_{12}	k% Vp	k% Ve	r_{12}	k% Vp	k% Ve	r_{12}	k% Vp	k% Ve
1	0,50	12	8	0,39	11	7	0,35	11	7
2	0,61	13	9	0,34	10	7	0,29	10	7
3	0,50	13	10	0,31	10	7	0,36	10	6
4	0,55	13	9	0,40	10	7	0,40	10	6
5	0,37	12	9	0,35	11	8	0,25	14	11
6	0,49	13	10	0,38	13	10	0,28	14	11
7	0,14	14	10	0,45	12	9	0,49	12	8
8	0,24	14	10	0,40	12	9	0,44	11	8
9	0,53	13	10	0,39	13	10	0,43	11	7
10	0,54	14	10	0,33	13	10	0,51	11	7
11	0,45	11	7	0,40	12	9	0,39	12	9
12	0,50	11	8	0,44	12	9	0,45	12	8
13	0,46	12	8	0,32	12	8	X	X	X
14	0,35	11	8	0,15	11	7	X	X	X
15	0,44	12	9	0,35	11	7	X	X	X
16	0,33	12	8	0,44	11	7	X	X	X

Pozn.: r_{12} = korelace faktorů 1 a 2; k%Vp = kumulativní podíl variance výsledků jednotlivých verzí testů, který je shrnutý oběma faktory před rotací (v procentech); k%Ve = kumulativní podíl variance výsledků jednotlivých verzí testů, který je shrnutý oběma faktory po rotaci (v procentech).

Jak lze z výše uvedené tabulky vidět, jsou ve všech verzích a ročnících SPF oba faktory mezi sebou slabě až středně silně korelovány. Zároveň nejsou jimi vysvětlené podíly variance příliš vysoké, zvláště po provedení rotace. I to může být příčinou velmi nízkých faktorových zátěží jednotlivých položek, které zřídka kdy překročily požadovanou hodnotu 0,3 (viz Urbánek, Denglerová, & Širůček, 2011). Příklad podoby těchto zátěží obsahuje Tab. VI v Příloze 5.

Z výše uvedených důvodů jsou výsledky faktorové analýzy v případě testů SPF napříč ročníky i verzemi jen obtížně interpretovatelné. Pro více informací jsme se opírali o výstupy z položkové analýzy (příklad podoby tohoto výstupu je k nalezení v Příloze 6). Při porovnání vlastností položek dle CTT, jejich faktorových zátěží v rotovaném modelu FA a

znění otázek v testu bylo zjištěno, že na jednom z faktorů mají častěji tendenci se shlukovat položky z oddílu logicko-analytického myšlení a položky cílící na prostorovou představivost, na druhém pak položky z verbálního oddílu týkající se slovní zásoby a otázky z literatury a historie z oddílu všeobecného přehledu. Při posuzování obtížnosti a diskriminační schopnosti těchto konkrétních položek nebylo shledáno, že by toto shlukování bylo dané právě jejich podobnou obtížností (ta se obvykle pohybovala v rozmezí 20-80%) či diskriminační schopností ($D = 0,2-0,4$ pro většinu položek). S opatrností by se tedy dalo hovořit o dvou potenciálních faktorech, které si pracovně dovolíme nazvat „logicko-analytickým myšlením a představivostí“ a „sečtělostí“.

Vzhledem k uvedeným problémům s interpretací výsledků faktorové analýzy jsme při vyhodnocování položkové analýzy jednotlivých verzí testu SPF v letech 2011 – 2013 nepřihlíželi ani tak k teoreticky definovaným oddílům (verbální myšlení, kritické myšlení, logicko-analytické myšlení, prostorová představivost, všeobecný přehled), jako spíše jsme se snažili o identifikaci konkrétních (typů) položek, které v daném roce a verzi často dosahovaly přiměřené obtížnosti (30-70%, s ideálem kolem 50%) a citlivosti ($D > 0,3$), případně se svými vlastnostmi těmto požadavkům konzistentně vymykaly. Takto jsme našli několik (byť velmi volných) zákonitostí:

- a) Položky zaměřené na *antonyma* byly většinou středně obtížné (45-70% správných odpovědí) a diferencovaly lépe ($D = 0,2-0,4$) než lehké (>80%) položky téhož druhu ($D = 0,14-0,25$).
- b) V SPF 2011 a 2012 byla *synonyma českých slov* buď výrazně obtížná (9-30%), ale diferencující ($D = 0,3-0,4$), nebo výrazně jednoduchá (>70%) a zároveň méně diferencující ($D < 0,25$). Naproti tomu obtížnější (<40%) *synonyma cizích slov* rozlišovala méně ($D < 0,2$) než jednoduchá (>70%) ($D = 0,27-0,38$). V SPF 2013 se již objevilo více položek středně obtížných (50-65%) a variabilnějších co do citlivosti v obou „okrajových“ skupinách (<30%; >70%), ač mnoho cizích synonym zůstalo velmi obtížných a špatně diferencujících.
- c) Otázky na *vztahy mezi slovy* („*X se má k Y jako Z k...*“) byly v letech 2011 a 2012 většinou výrazně jednoduché (>90%) s nižší citlivostí ($D < 0,2$), v SPF 2013 se již začaly objevovat položky spíše středně obtížné (50-70%), diferencující lépe ($D = 0,2-0,37$).

- d) Otázky na význam frází a archaismy jsou vesměs velmi variabilní svou obtížností i citlivostí (napříč celým spektrem hodnot), středně obtížné (50-70%) položky ovšem diferencují lépe než položky snadné (>80%) a příliš těžké (<30%).
- e) Otázky z oblasti *kritického myšlení* byly v SPF 2011 spíše středně obtížné až snadné (50-85%), v SPF 2012 se obtížnost položek zvýšila (30-60%). V testech SPF 2013 byl počet otázek středně těžkých (50-60%), těžkých (30-40%) a snadných (>80%) poměrně vyrovnaný. Zlepšila se však citlivost snazších položek, které v předchozích letech stejně jako položky příliš obtížné diferencovaly méně ($D < 0,2$).
- f) *Doplňování řad* písmen, čísel či obrázků je ve většině případů snadné (>80%) a nepříliš diferencující ($D < 0,2$). Lepší rozlišovací účinnost ($D > 0,25$) vykazují položky zaměřené na *doplňování čísel do obrazců* (trojúhelníků, čtyřúhelníků), i tyto položky jsou ovšem převážně velmi lehké (>80%).
- g) *Slovní úlohy* (fyzikální, nepřímé úměrnosti, počítající s průměry) byly téměř bez výjimky obtížné až středně těžké (<50%) s velmi dobrou rozlišovací schopností ($D > 0,3$). Pouze úlohy počítající s procenty dosahovaly při stejné obtížnosti variabilní citlivosti.
- h) Otázky pracující s *výrokovou logikou* byly v SPF 2011 většinou lehké (>80%), ale uspokojivě rozlišující ($D = 0,25-0,3$). V SPF 2012 jejich obtížnost stoupla (40-60%), některé výroky se však staly spíše „tipovacími“ ($D < 0,1$), zatímco u jiných diskriminační síla přetrvala.
- i) Při *skládání obrazců* hraje velkou roli typ úlohy; skládání „rozložené krychle“ se většinou ukázalo jako středně obtížné až lehké (50-80%), ale velmi dobře diferencující ($D > 0,25$), zatímco skládání osmistěny bylo bez výjimky vysoce obtížné (<30%) a ztrácelo rozlišovací schopnost ($D < 0,15$). Skládání písmen z jejich částí je převážně středně obtížné až snadné (50-80%), rozlišovací schopnost této úlohy je však spíše nižší ($D < 0,2$).
- j) *Dvojdímenzionální rotace QR-kódů* vykazovala bez výjimky střední obtížnost (40-50%) a velmi dobré diskriminační schopnosti ($D > 0,3$). Naproti tomu rotace 3D objektů byla spíše středně obtížná až lehká (50-80%) a její citlivost kolísala okolo hodnoty $D = 0,3$.

- k) Úlohy pracující s *chybějícími kostkami v krychli* byly ve verzích SPF 2012 konstantně středně obtížné (40-60%) s vysokou rozlišovací účinností ($D > 0,4$). Ta zůstala převážně zachována i v SPF 2013 ($D = 0,3-0,4$), ačkoliv úlohy zde byly spíše středně obtížné až snadné (50-80%).
- l) Otázky *všeobecného přehledu* byly ve všech ročnících spíše obtížnější až průměrně obtížné (30-60%) s variabilní citlivostí. Vzhledem k jejich obsahové nesourodosti nebylo možné dohledat nějaký „spojující“ prvek či vzorec, s ohledem na jejich počet se však v této sekci častěji vyskytovaly „vadné“ položky, u nichž celkově méně úspěšní uchazeči odpovídali správně častěji než 27% nejuspěšnějších uchazečů.

Celkově lze shrnout, že obsahově je test SPF velmi variabilní jak z hlediska obtížnosti položek, tak z hlediska jejich rozlišovací účinnosti. Tato variabilita přitom prochází nejen napříč testem, ale i jeho jednotlivými (teoretickými) oddíly, a patrně se odráží i v interpretačně „slabých“ výsledcích explorační faktorové analýzy.

9.3 Výsledky ověřování prediktivní a inkrementální validity testů SPF

Hodnota jednoduché Pearsonovy korelace mezi body SPF a známkovým průměrem v prvním roce studia činila v akad. roce 2011/2012 $r_{(1223)} = -0,09$ ($p = 0,003$). Po korekci pro omezení variability se hodnota zvedla na $r_{(1223)} = -0,10$. Pro akad. rok 2012/2013 činila nekorigovaná Pearsonova korelace mezi body SPF a známkami v prvním ročníku $r_{(1318)} = -0,24$ ($p < 0,001$), po korekci pak $r_{(1318)} = -0,26$. Výsledek v testu SPF tedy dovede samostatně vysvětlit 1-7% variance kritéria. Na hladině $\alpha = 0,05$ zároveň **přijímáme hypotézy H4a a H4b**. Skór testu SPF v roce 2011 i v roce 2012 statisticky významně koreluje s průměrem známek studentů v prvním ročníku.

Pearsonova korelace mezi maturitním průměrem a známkami v prvním roce studia na VŠ činila v akad. roce 2011/2012 $r_{(821)} = 0,25$ ($p < 0,001$), v akad. roce 2012/2013 pak $r_{(973)} = 0,29$ ($p < 0,001$). Maturitní průměr tedy statisticky významně koreluje se známkami v prvním roce studia na VŠ a samostatně vysvětluje 6-8% variance tohoto kritéria.

Pearsonova korelace mezi věkem a známkovým průměrem v prvním ročníku činila v akad. roce 2011/2012 $r_{(1250)} = -0,14$ ($p < 0,001$), v akad. roce 2012/2013 pak $r_{(1331)} = 0$. Věk samostatně tedy v celkovém souboru zapsaných uchazečů se známkovým průměrem nekoreluje a samostatně objasňuje 0-2% variance kritéria.

Rozdíl mezi známkovými průměry na konci prvního ročníku u uchazečů přijatých v akad. roce 2011/2012 a 2012/2013 na základě přijímací zkoušky, odvolání a bez přijímací zkoušky shrnuje Tab.8. Leveneho test pro homogenitu rozptylů odhalil v akad. roce 2011/2012 signifikantní rozdíl mezi rozptyly jednotlivých skupin ($F_{(2,1247)} = 12,99$; $p < 0,001$), proto byl pro ověření rozdílů ve známkových průměrech v tomto akademickém roce použit Brown-Forsythův test, který je robustní variantou jednofaktorové analýzy variance.

Tab. 8: Výsledky jednofaktorové analýzy variance a Brown-Forsythova testu

	N	Pr.Z.	SD	F	sv1	sv2	p
Ak. rok 2011/2012							
s PZk	1.207	2,41	0,92	9,13	2	21,44	0,001
bez PZk	27	2,01	0,34				
odvolání	16	2,87	0,85				
Ak. rok 2011/2012							
s PZk	1.310	2,54	0,92	1,84	2	1328	0,16
bez PZk	13	2,07	0,63				
odvolání	8	2,37	1,01				

Pozn.: Pr.Z. = průměr známek v prvním ročníku studia; SD = standardní odchylka; sv1 a sv2 = stupně volnosti.

Mezi třemi skupinami uchazečů přijatých v akad. roce 2011/2012 byl nalezen statisticky významný rozdíl ve známkových průměrech na konci prvního ročníku. Uchazeči přijatí bez přijímací zkoušky vykazují nižší známkový průměr než uchazeči přijatí po absolvování PZk a uchazeči přijatí na odvolání. Jak vyplývá z post-hoc testů dle Dunettova T3 pro nehomogenní variance, jsou tyto rozdíly statisticky významné ($p < 0,001$ pro skupiny „s PZk“ a „bez PZk“; $p = 0,003$ pro skupiny „odvolání“ a „bez PZk“). Podobný trend je patrný i u uchazečů v akad. roce 2012/2013, zde však rozdíly mezi skupinami nedosahují statistické významnosti.

Výsledky krokové regresní analýzy pro ověření hypotéz H5a-b jsou shrnuty v Tab.9. Příloha 7 ještě obsahuje tabulku s jednotlivými regresními koeficienty a výsledky jednofaktorové analýzy rozptylu pro vhodnost daného modelu.

Tab. 9: Přírůstek R^2 v krokové regresní analýze pro ověření H5a a H5b

	R	R^2	kor. R^2	SDCh	přír. R^2	F	sv1	sv2	p
Akad. rok 2011/2012									
1 (mat. průměr)	0,252	0,063	0,062	0,50	,063	53,889	1	795	0,000
2 (mat. průměr, body SPF)	0,267	0,071	0,069	0,50	,008	6,792	1	794	0,009
3 (mat. průměr, body SPF, věk)	0,393	0,154	0,151	0,47	,083	77,758	1	793	0,000
Akad. rok 2012/2013									
1 (mat. průměr)	0,294	0,087	0,086	0,61	,087	90,661	1	957	0,000
2 (mat. průměr, body SPF)	0,367	0,134	0,133	0,59	,048	52,937	1	956	0,000
3 (mat. průměr, body SPF, věk)	0,368	0,135	0,132	0,59	,001	,665	1	955	0,415

Pozn.: kor. R^2 = korigovaná hodnota determinačního koeficientu (populační odhad) ; přír. R^2 = přírůstek v koeficientu determinance; SDCh = standardní chyba odhadu; sv1 a sv2 = stupně volnosti.

Přírůstek podílu variance R^2 ve druhém kroku, kdy byl zadán skór testu SPF, je v obou letech (akad. rok 2011/2012 a 2012/2013) statisticky významný. **Přijímáme tedy hypotézy H5a a H5b.** Výsledek testu SPF dovede predikovat známky v prvním roce studia i nad rámec maturitního průměru studentů. Tento efekt je o něco výraznější v akad. roce 2012/2013 (přírůstek $R^2 = 0,048$). V tomto roce se také neukázal být významným přínos věku nad rámec zbylých prediktorů, přičemž v akad. roce 2011/2012 dovedl věk objasnit dalších 8% variance. Celkový podíl variance kritéria vysvětlený modelem se všemi prediktory činil pro rok 2011/2012 přibližně 15%, pro rok 2012/2013 přibližně 13%. Podotkněme ještě, že při pohledu na standardizované beta-koeficienty jednotlivých prediktorů (Příloha 7) má v obou letech maturitní průměr v absolutních hodnotách vyšší váhu než test SPF, přičemž tento rozdíl je výraznější v akad. roce 2011/2012.

Tabulka 10 shrnuje výsledky dalšího modelu regresní analýzy, tentokrát pro ověření hypotéz H6a a H6c – přínos testu SPF nad rámec „zbylé části zkoušky“ v oborech s váhou SPF 30%. Údaje pro obor psychologie v akad. roce 2011/2012 (H6b, váha SPF 33%) jsou uvedeny v Tab. 11. Jednotlivé váhy regresních koeficientů lze, opět, dohledat v tabulkách v Příloze 8 a Příloze 9.

Tab. 10: Přírůstek R^2 v krokové regresní analýze pro ověření H6a a H6c (váha SPF 30%)

	R	R^2	kor. R^2	SDCh	přír. R^2	F	sv1	sv2	p
Ak. rok 2011/2012									
1 (OT/ÚZ aj.)	0,051	0,003	0,000	0,51	,003	0,889	1	343	0,346
2 (OT/ÚZ aj., body SPF)	0,053	0,003	-0,003	0,51	,000	0,059	1	342	0,808
3 (OT/ÚZ aj., body SPF, mat. průměr)	0,146	0,021	0,013	0,50	,018	6,415	1	341	0,012
Ak. rok 2012/2013									
1 (OT/ÚZ aj.)	0,390	0,152	0,150	0,60	,152	77,135	1	431	0,000
2 (OT/ÚZ aj., body SPF)	0,546	0,299	0,295	0,54	,147	90,000	1	430	0,000
3 (OT/ÚZ aj., body SPF, mat. průměr)	0,564	0,318	0,314	0,54	,020	12,523	1	429	0,000

Pozn.: kor. R^2 = korigovaná hodnota determinačního koeficientu (populační odhad) ; přír. R^2 = přírůstek v koeficientu determinance; SDCh = standardní chyba odhadu; sv1 a sv2 = stupně volnosti.

Tab. 11: Přírůstek R^2 v krokové regresní analýze pro ověření H6b (váha SPF 33%)

Krok	R	R^2	kor. R^2	SDCh	přír. R^2	F	sv1	sv2	p
1 (OT/ÚZ aj.)	0,004	0,000	-0,011	0,35	,000	0,001	1	88	0,971
2 (OT/ÚZ aj., body SPF)	0,057	0,003	-0,020	0,35	,003	0,286	1	87	0,594
3 (OT/ÚZ aj., body SPF, mat. průměr)	0,345	0,119	0,088	0,33	,116	11,318	1	86	0,001

Pozn.: kor. R^2 = korigovaná hodnota determinačního koeficientu (populační odhad) ; přír. R^2 = přírůstek v koeficientu determinance; SDCh = standardní chyba odhadu; sv1 a sv2 = stupně volnosti.

Z uvedených tabulek vyplývá, že v akad. roce 2011/2012 *není* přírůstek R^2 po zadání výsledku testu SPF nad rámec zbylé části přijímací zkoušky statisticky významný, a to jak v oborech s váhou SPF 30%, tak ani v oboru psychologie. **Zamítáme tedy hypotézy H6a a H6b.** Výsledek testu SPF v tomto akademickém roce nevysvětluje u oborů s dalšími povinnými součástmi PZk prakticky žádnou variaci v kritériu. Stejně tak tomu však je u zbylé části PZk. Jediným signifikantním prediktorem známkového průměru v prvním ročníku studia je tak v akad. roce 2011/2012 maturitní průměr, jenž i nad rámec celé přijímací zkoušky objasňuje při váze testu SPF 30% asi 2% variance, v oboru psychologie pak asi 11-12% variance. Při pohledu na výsledky jednofaktorové analýzy variance (Příloha 8 a 9) shledáme, že při váze testu SPF 30% není v akad. roce 2011/2012 ani model

se všemi prediktory statisticky významný ($F_{(3, 341)} = 2,459$; $p = 0,06$). U oboru psychologie již celkový model stat. významnosti dosahuje ($F_{(3, 86)} = 3,88$; $p = 0,01$).

V akad. roce 2012/2013 již část přijímací zkoušky mimo test SPF vysvětluje 15% variance a je statisticky významná. Stejně tak je ovšem významný i přírůstek $R^2 = 0,147$ po zadání výsledku testu SPF do regresní rovnice. **Přijímáme tedy hypotézu H6c.** Výsledek testu SPF (s váhou 30%) vysvětluje v tomto akad. roce cca 15% variance nad rámec zbylé části přijímací zkoušky a při pohledu na standardizované beta-koeficienty (Příloha 8) zjistíme, že je v kompletním modelu silnějším prediktorem než zbylá část PZk. Také přírůstek $R^2 = 0,02$ po zadání maturitních průměrů je statisticky významný a je srovnatelný s efektem tohoto prediktoru v roce 2011/2012. Kompletní model pak v tomto akad. roce vysvětluje asi 31-32% variance.

Výsledky potřebné pro ověření poslední sady hypotéz (H7a-c), totiž zda je test SPF přínosný i nad rámec znalosti maturitního průměru a výsledku u oborového testu zároveň, jsou uvedené v Tab. 12 a 13; pro beta-koeficienty i zde odkazujeme na Přílohy 9 a 10.

Tab. 12: Přírůstek R^2 v krokové regresní analýze pro ověření H7a a H7c (váha SPF 30%)

	R	R^2	kor. R^2	SDCh	přír. R^2	F	sv1	sv2	p
Ak. rok 2011/2012									
1 (mat.průměr)	0,125	0,016	0,013	0,50	,016	5,485	1	343	0,020
2 (mat.průměr, OT/ÚZ aj.)	0,143	0,021	0,015	0,50	,005	1,672	1	342	0,197
3 (mat.průměr, OT/ÚZ aj., body SPF)	0,146	0,021	0,013	0,50	,001	0,224	1	341	0,636
4 (mat.průměr, OT/ÚZ aj., body SPF, věk)	0,420	0,177	0,167	0,46	,155	64,178	1	340	0,000
Ak. rok 2012/2013									
1 (mat.průměr)	0,313	0,098	0,096	0,61	,098	46,694	1	431	0,000
2 (mat.průměr, OT/ÚZ aj.)	0,448	0,201	0,197	0,58	,103	55,592	1	430	0,000
3 (mat.průměr, OT/ÚZ aj., body SPF)	0,564	0,318	0,314	0,54	,117	73,938	1	429	0,000
4 (mat.průměr, OT/ÚZ aj., body SPF, věk)	0,580	0,336	0,330	0,53	,017	11,193	1	428	0,001

Pozn.: kor. R^2 = korigovaná hodnota determinačního koeficientu (populační odhad) ; přír. R^2 = přírůstek v koeficientu determinance; SDCh = standardní chyba odhadu; sv1 a sv2 = stupně volnosti.

Z uvedené tabulky vyplývá, že u oborů s váhou SPF 30% jsou v akad. roce 2011/2012 jedinými signifikantními prediktory maturitní průměr, vysvětlující samostatně opět 1-2% variance kritéria, a věk, a to i nad rámec ostatních proměnných (přírůstek $R^2 = 0,155$; $p < 0,001$). Po zadání výsledků zbylé části přijímací zkoušky a skóru v testu SPF nad rámec maturitního průměru se podíl vysvětlené variance ani v jednom případě významně nezvýší. **Zamítáme tedy hypotézu H7a.** Celkově vysvětluje tento model v akad. roce 2011/2012 asi 18% variance, přičemž věk je dokonce lepším prediktorem než mat. průměr (viz Příloha 10).

Tab. 13: Přírůstek R^2 v krokové regresní analýze pro ověření H7b (váha SPF 33%)

Krok	R	R^2	kor. R^2	SDCh	přír. R^2	F	sv1	sv2	p
1 (mat.průměr)	0,343	0,118	0,108	0,34	,118	11,736	1	88	0,001
2 (mat.průměr, OT/ÚZ aj.)	0,343	0,118	0,098	0,33	,000	0,013	1	87	0,908
3 (mat.průměr, OT/ÚZ aj., body SPF)	0,345	0,119	0,088	0,33	,001	0,137	1	86	0,712
4 (mat.průměr, OT/ÚZ aj., body SPF, věk)	0,393	0,155	0,115	0,32	,036	3,579	1	85	0,062

Pozn.: kor. R^2 =korigovaná hodnota determinačního koeficientu (populační odhad) ; přír. R^2 = přírůstek v koeficientu determinance; SDCh = standardní chyba odhadu; sv1 a sv2 = stupně volnosti.

Pro studenty oboru psychologie v akad. roce 2011/2012 (váha SPF 33%) lze z Tab. 13 vyčíst, že ani zde se R^2 po zadání výsledků z testu SPF nad rámec maturitního průměru a bodů získaných ve zbylé části PZk statisticky významně nezvýší. **Zamítáme tedy hypotézu H7b.** Zároveň je u této skupiny studentů velmi výrazným prediktorem maturitní průměr, vysvětlující 11% variance (z pohledu do Tab.X v Příloze 9 je patrné, že maturitní průměr je nejsilnějším prediktorem i v celkovém modelu regresní analýzy). Naopak přínos věku $R^2 = 0,036$ nad rámec ostatních prediktorů přestává být statisticky významný ($p = 0,65$). Celkový model se čtyřmi prediktory a váhou SPF 33% dohromady vysvětluje asi 15% variance kritéria.

V akad. roce 2012/2013 se u analyzovaných oborů s váhou SPF 30% jeví statisticky významným každý prediktor zadaný do regresní analýzy. Maturitní průměr samostatně vysvětluje asi 10% variance kritéria, dalších 10% pak vysvětluje část přijímací zkoušky mimo SPF. Po zadání skóru SPF do rovnice pozorujeme přírůstek $R^2 = 0,117$ (tzn. asi 12% variance vysvětlené SPF nad rámec mat. průměru a zbylé části PZk). **Přijímáme**

tedy hypotézu H7c. Efekt věku na predikci známkového průměru je v akad. roce 2012/2013 nad rámec ostatních prediktorů sice statisticky významný, v absolutních hodnotách se však jedná o cca 2% vysvětlené variance navíc. Celkový model v tomto roce pak objasňuje asi 34% variance kritéria. Z pohledu na standardizované beta-koefficienty (Příloha 10) v kompletním modelu ještě shledáme, že vliv maturitního průměru na známky v prvním roce studia se po zadání výsledku v testu SPF výrazně sníží. Test SPF je také v absolutních hodnotách standardizovaných beta-koefficientů nejsilnějším ze čtyř ověřovaných prediktorů, následovaný výsledkem ve zbylé části PZk.

Souhrnně lze říci, že predikční schopnost testu SPF nad rámec dalších prediktorů byla v uplynulých letech variabilní, závislá i na počtu a druhu srovnávaných proměnných, resp. zkoumaném vzorku. Podíl variance vysvětlené testem SPF samostatně i nad rámec ostatních proměnných v regresní analýze se pohyboval v rozmezí 0-12%. Poměrně výrazným byl v akad. roce 2011/2012 vliv věku, který se v analýzách pro rok 2012/2013 ztrácel. Naopak maturitní průměr dovedl poměrně stabilně predikovat známkový průměr v prvním ročníku; samostatně vysvětloval asi 6-8% variance kritéria, nad rámec ostatních prediktorů pak nejčastěji cca 2% variance kritéria. Výjimkou byla skupina studentů oboru psychologie v akad. roce 2011/2012, u nichž maturitní průměr samostatně i nad rámec ostatních proměnných vysvětloval asi 12% variance. Kompletní modely pak celkově objasňovaly 2-32% variance známkového průměru v prvním roce studia na VŠ.

10 Diskuze

Při sestavování výzkumné části této práce jsme narazili na několik metodologických obtíží s ověřováním reliability a validity testů SPF podle v literatuře popisovaných postupů, které bychom na tomto místě rádi dále rozvedli a zasadili do kontextu našich výsledků.

V první řadě se jednalo o nemožnost posouzení paralelity jednotlivých verzí testu (v jednotlivých ročnících, případně i napříč ročníky) na základě srovnání dat týchž uchazečů. K tomu by bylo zapotřebí, aby každý z nich absolvoval alespoň dvě, ne-li pak všechny verze testu SPF, a jejich výsledky by byly korelovány. Tento postup není možné v rámci přijímacího řízení na FF UP aplikovat, nabízí se však možnost pilotního ověření jednotlivých verzí testu na dostatečně velkém vzorku reprezentativním pro cílovou populaci. Poté by mohla být srovnána paralelita jednotlivých verzí a do „ostrého“ testu by mohly být zařazeny již pouze položky, které vykazují dobré psychometrické vlastnosti, čímž by se kvalita tohoto testu zvýšila. Problémů spojených s pilotním testováním je však také několik: Zaprvé by bylo nutné nejprve přesně definovat „cílovou populaci“ testů SPF; je otázka, zda by se jednalo o všechny občany ČR přiměřeného věku, kteří by se mohli na VŠ hlásit (jak by ale do této definice zapadali uchazeči např. ze Slovenska?), nebo by specifikace byla užší – a pokud ano, v jakém ohledu a proč. Zadruhé by tento postup výrazně navýšil rozpočet na přijímací zkoušky, a to jak finančně, tak personálně a časově. Bylo by potřeba vyvinout mnohem více testových otázek, neboť z praxe je známo, že jejich počet se po provedení položkové analýzy (a ověření jejich vlastností) výrazně redukuje. Je otázkou, zda by FF UP byla ochotna či vůbec disponovala prostředky na to, aby k takovému navýšení přistoupit. Největší starostí by však bylo riziko prozrazení otázek „skutečným“ uchazečům, které by (i přes případné smluvní závazky pilotních probandů o mlčenlivosti) nebylo fakticky možné ohlídat. S ohledem na tyto skutečnosti jsme tedy mohli pracovat pouze s daty reálných účastníků přijímacího řízení na FF UP a testu SPF v té podobě, v jaké nyní jsou, tj. od každého uchazeče byly k dispozici pouze jeho výsledky z jediné absolvované verze testu. S odkazem na odbornou literaturu (např. Urbánek, Denglerová, & Širůček, 2011), která udává, že při vzájemné paralelitě verzí testu jsou průměrné výsledky (a jejich směrodatné odchylky) uchazečů ve všech verzích srovnatelné jsme se rozhodli přistoupit k ověření alespoň tohoto předpokladu a provedli sadu dvouvýběrových t-testů a F-testů pro srovnání rozptylů. Protože však literatura počítá s tím, že se i v případě těchto srovnání bude jednat o data *týchž* uchazečů, je potřeba naše

výsledky interpretovat s opatrností. Již v teoretické části této práce jsme zmínili, že při administraci testů SPF je problematická už její standardizace, totiž zachování stejných pracovních podmínek pro všechny uchazeče. Z kapacitních důvodů učeben na FF UP není možné, aby všichni uchazeči absolvovali test naráz a v jedné (standardně vybavené) místnosti, a tak testování probíhá ve více učebnách a „po etapách“. Vzhledem k výše uvedenému riziku prozrazení testových otázek pak ovšem jednotlivé „várky“ uchazečů nedostávají tytéž verze testu či jejich kombinace (viz dále). Tento fakt se dále mísí se skutečností, že uchazeči v různých várkách se většinou hlásí na různé skupiny oborů, neboť někteří krom testu SPF absolvují v daný den ještě oborový test a je logisticky snazší seztvat uchazeče o stejný obor na stejný čas a místo. V samotných výsledcích různých verzí testů SPF tak ovšem mohou vznikat systematické chyby, neboť je možné, že tyto „jinak namíchané“ várky už předem vykazují rozdíly v námi ověřovaných schopnostech (např. uchazeči se zájmem o muzikologii nemusí disponovat stejnými charakteristikami jako uchazeči se zájmem o angličtinu a aplikovanou ekonomii)³⁵. Jelikož jsou z hlediska další úspory v každé várce dány do oběhu „párové“ verze SPF (tzn. obsahující shodné otázky, pouze v jiném pořadí či s jinak označenou správnou odpovědí), je de facto možné porovnávat pouze jejich parametry – tedy parametry každého páru (liché a sude verze) – zvláště, neboť pouze ty jsou ověřovány na „stejně“ sub-populaci. Jelikož při mnohonásobných porovnáváních stoupá statistická pravděpodobnost nalezení falešně pozitivního výsledku, bylo třeba korigovat hodnotu alfa na hladinu 0,006 pro verze SPF 2011 a 2012 (16 verzí celkem, tzn. 8 porovnání) a na hladinu 0,008 pro verze SPF 2013 (12 verzí celkem, tzn. 6 porovnání). Za těchto podmínek se sice neukázal statisticky významný rozdíl v průměrech a směrodatných odchylkách jednotlivých verzí, na hladině alfa = 0,05 by se však od sebe některé varianty lišily. Není ovšem možné přesně říci, zda proto, že byly skutečně jinak obtížné než varianty jiné, nebo proto, že ta část studentů, která tyto varianty absolvovala, se svými vlastnostmi odlišovala od studentů v jiných várkách. Stejnou chybou může být zatížení srovnání vnitřních konzistencí párových verzí testu, k němuž jsme přistoupili ze stejné logiky jako ke srovnání průměrů a směrodatných odchylek, totiž že testy, které obsahují stejné otázky a jsou administrované stejné sub-skupině uchazečů, by měly také vykazovat stejné hodnoty vnitřní konzistence. Ty se v jednotlivých ročnících a verzích testů SPF pohybovaly v rozmezí $r = 0,6-0,77$, což lze považovat spíše za spodní hranice dobré vnitřní konzistence doporučené pro

³⁵ Je také otázkou, nakolik vlastně chceme či máme chtít, aby si tito uchazeči byli svými vlastnostmi, schopnostmi a dovednostmi podobní, o tom však na jiném místě.

psychodiagnostické testy v odborné literatuře (např. Schultz, & Whitney, 2005). Jedná se také o něco slabší výsledky, než jakých dosahují testy TSP MU v Brně (viz Žoudlík, 2009); možným příčinám této nižší homogenity se budeme věnovat dále v rozboru obsahové validity SPF. Mezi vnitřními konzistencemi lichých a sudých verzí každého páru také nebyl nalezen statisticky významný rozdíl, jako v předchozích případech však nelze vyvozovat, že verze jsou si tím pádem paralelní i navzájem, totiž „napříč“ páry. Do budoucna bychom doporučovali zvážit možnost pilotního ověřování otázek a navýšení rozpočtu na jejich tvorbu; pokud by se nám totiž podařilo vyvinout dostatečně rozsáhlý soubor otázek (stovky, či spíše tisícovky položek, ideálně v odpovídající kvalitě), mělo by teoreticky být riziko zkreslení skutečných výsledků uchazečů jejich prozrazením klesnout, neboť nepředpokládáme, že by se uchazeči byli schopni naučit nazpaměť všechny, i kdyby jejich znění znali předem.

Pokud jde o validitu samotného obsahu testů SPF, její posuzování komplikuje fakt, že konstrukt „studijních předpokladů“ není v literatuře jasně definován a jeho podoba, náplň či odlišnost od konceptu obecné inteligence zůstává svým způsobem nedořešena (viz např. Gottfredsonová, 2004). Z tohoto důvodu je složité již vůbec zhodnocení obsahové validity testu SPF expertním odhadem (které by mělo probíhat souběžně s jeho tvorbou) či za pomoci content validity ratio, neboť názory odborníků na to, co do testu má či nemá patřit, se mohou značně různit. Většina testů studijních předpokladů se sice zaměřuje na „schopnost uvažovat“ v oblastech jako je kvantitativní či analytické myšlení, verbální a kritické myšlení, abstraktní myšlení nebo prostorová představivost aj. (např. Lawrence, Rigol, Van Essen, & Jackson, 2002; Masarykova univerzita v Brně, 20. května 2013; Scio, 20. května 2013), není však řečeno, zda jde o absolutní výčet dimenzí studijních předpokladů. Otázka dimenzionality je přitom pro tvorbu a ověřování psychometrických vlastností testů zásadní (např. Urbánek, Denglerová, & Širůček, 2011). V naší práci jsme za pomoci explorační faktorové analýzy s neortogonální rotací zjistili, že v našich datech se vyskytuje mnoho náhodné či residuální variance, kterou nebylo schopno postihnout ani prvotně vyextrahovaných 19 – 22 faktorů s vlastním číslem vyšším než 1 (Kaiserovo kritérium). Po prostudování sutinových grafů jsme navíc usoudili, že nemá smysl z dat extrahovat a dále rotovat více jak dva faktory, neboť přírůstky vysvětlené variance byly jen velmi nízké. I tyto dva faktory ovšem dohromady objasňovaly nejvýše 14% variance a faktorové zátěže většiny položek jen stěží dosahovaly hodnoty 0,3, která je doporučovaná jako minimální mez k tvrzení, že daná položka spadá do dané dimenze (Urbánek,

Denglerová, & Širůček, 2011). Položky určitého typu (logicko-analytické myšlení, prostorová představivost, položky spojené se slovní zásobou a znalostmi literatury a historie) sice měly tendenci se napříč verzemi shlukovat do jednotlivých faktorů, tato tendence však byla pouze velmi slabá. Naše pracovní interpretace, že by se nalezené faktory daly označit jako „logicko-analytické myšlení a představivost“ a „sečtělost“, lze tedy interpretovat jen s nejvyšší opatrností. Zůstává pro nás otázkou, co variabilitu v získaných datech způsobilo. Jednou z možností je malý vzorek uchazečů, neboť jejich počty v některých verzích skutečně nedosahovaly ani minimální hranice $n = 100$, natož pak doporučeného poměru k počtu položek $n:k = 5:1$, tzn. $n > 250$. To se odrazilo i v hodnotách Kaiser-Meyer-Olkinova koeficientu, který byl obecně spíše slabší až středně silný. Vysokou variabilitu ovšem vykazovaly i verze, kde byla obě kritéria splněna, tzn. provedení faktorové analýzy by mělo mít smysl. Je ovšem také možné, že testy SPF jsou de facto směsí znalostních otázek (zvláště s ohledem na vysoký podíl otázek k všeobecného přehledu) a otázek kopírujících obsahy zavedených IQ testů, přičemž právě z tohoto důvodu vykazují vyšší heterogenitu, která se ostatně projevila i v koeficientech vnitřní konzistence. (Je možné se tázat, zda i z tohoto důvodu poté nebyla predikční schopnost testů SPF vůči známám v prvním ročníku studia slabší, a to především v roce 2011; více viz níže.)

Také jednotlivé položky samotné jsou různě obtížné a vykazují různou schopnost rozlišovat mezi úspěšnými a méně úspěšnými uchazeči, a to jak vůči sobě navzájem, tak v rámci jednoho typu položek. Takto jsme například zjistili, že otázky v (zamýšlené) části verbálního myšlení jsou vždy tak trochu sázkou do loterie, i když synonyma méně známých slov domácího původu (případně jejich antonyma) a synonyma známějších cizích slov by mohla být do testů zařazena. V (teoretické) logicko-analytické části bychom doporučili pokračovat v typově podobných slovních úlohách jako v SPF 2013 (nepřímé úměry), neboť dosahují střední obtížnosti a uspokojivé citlivosti. Možné by bylo doplnit je úlohami na skládání pláštů krychlí, které je sice středně obtížné až snadné, ale vykazuje dobrou diskriminační schopnost. Naopak skládání osmistěnnů se ukázalo jako konstantně velmi obtížné a nerozlišující a z hlediska obtížnosti a rozlišovací schopnosti jsou slabé i položky na skládání písmen z jejich částí. Místo doplňování řad čísel, písmen a obrázků by pak bylo lépe zvýšit počet úloh s doplňováním čísel do obrazců, ačkoliv i tyto položky jsou poměrně lehké. Určitě bychom ale doporučovali zachovat či opět zavést rotace QR kódů či podobně komplikovaných 2D objektů v části prostorové představivosti, neboť tyto položky

mají ideální střední obtížnost při současné vynikající rozlišovací schopnosti. Podobně vhodné jsou také úlohy s doplňováním kostek do krychle.

U položek ze všeobecného přehledu je stále k zamyšlení, zda je v testu ponechat či nikoliv. Jejich obtížnost byla většinou vyšší (30-60% správných odpovědí celkem), často se ovšem stávalo, že ztrácely rozlišovací schopnost, případně se mezi nimi objevovaly položky, na které 27% nejméně úspěšných uchazečů odpovídalo správně častěji než 27% těch nejméně úspěšných. Je otázkou, zda byly distraktory v těchto položkách natolik kvalitní, že se jinak úspěšní uchazeči nechali zmást (příliš nad nimi přemýšleli), nebo zda šlo o náhodný výsledek způsobený tím, že odpovědi na obtížné položky uchazeči „tipovali“. Žoudlík (2009) v tomto ohledu diskutuje možnost penalizace za špatné odpovědi, čímž by se tipování mělo snížit, Crocker a Algina (2008) ovšem upozorňují, že při možnosti strhávání bodů dochází k dalším komplikacím při výpočtu reliability testu. Z hlediska obsahové validity je ovšem obtížné vůbec určit, které znalosti jsou natolik „základní“ či „nutné“, že je musí každý uchazeč o vysokoškolské studium bezpodmínečně znát. Toto rozhodnutí je často arbitrární a závislé na autorech testu.

Z druhé strany může být vysoký rozptyl v datech způsobený i samotným vzorkem uchazečů, na kterých testování probíhalo. Crocker a Algina (2008) upozorňují, že není-li soubor testovaných dostatečně variabilní co do pravých skóre, většina testové variance bude připsána chybovému rozptylu. To může i důvod vysoké residuální variance v našich datech. Je - i vzhledem k nejasně definované populaci uchazečů o studium na VŠ - možné, že lidé, kteří se na VŠ hlásí, již předem vykazují lepší schopnosti (které je následně k tomuto rozhodnutí vedly) než ti, kteří o VŠ studium neusilují.

Omezení variability se dále mohlo projevit ve výpočtech prediktivní validity testu SPF a dalších proměnných vůči známkovému průměru (přijatých) uchazečů v prvním roce studia. Ten je spolu s celkovým známkovým průměrem (GPA) ve validizačních studiích testů studijních předpokladů v zahraničí nejčastějším kritériem akademického úspěchu (např. Camara, 2004a). Námi nalezené hodnoty jednoduché Pearsonovy korelace mezi skórem testu SPF a průměrem známek na konci prvního ročníku však byly pouze velmi slabé ($r = -0,09$ pro SPF 2011 a $r = -0,24$ pro SPF 2012). Zahraniční studie testů SAT dle Camary (2004a) většinou nacházejí korelace v rozmezí přibližně $r = 0,44-0,62$. Burton a Ramist (2001) udávají, že průměrná korelace mezi skórem SAT a GPA v odborných studiích činí průměrně $r = 0,36$ (s rozpětím 0,22-0,52), pro SAT a známky v prvním roce

studia potom $r = 0,35$. Z českých výzkumů lze ovšem pro naše účely považovat za srovnatelné výsledky analýzy testu VSP na 1.LF UK. Štuka (2012) udává, že jím nalezená korelace celkového skóru z odborných testů se známkovým průměrem studentů medicíny v prvním roce byla vyšší ($r = 0,3$) než korelace testu VSP ($r = 0,13$). Stejně jako v našem případě se jedná o poměrně nízké hodnoty. Je však pravděpodobné, že oba výpočty trpí omezením variability, a to z dvojího hlediska: Jednak už samotné zvolené kritérium, totiž známky na VŠ, je svými vlastnostmi značně problematické. Při standardním známkování A-F, které de facto odpovídá známkám 1-4, může velmi snadno dojít k „efektu stropu“, totiž hromadění známek na jednom konci (už tak poměrně úzkého) spektra (Říčan, 1977). To pak obratem může zkreslit predikční schopnost testů vzhledem ke známkám v prvním a dalších ročnících VŠ – kde není variabilita, není reliabilita ani validita. Zároveň Laird (2004) uvádí, že známkování (nejen na VŠ, ale v podstatě na jakékoliv škole) není zcela reliabilním kritériem a zvláště v humanitních oborech může být velmi mírné. Obzvláště tehdy, pokud jsou brány v potaz známky z různých předmětů a od různých vyučujících, jako tomu bylo v našem případě, je možné očekávat, že toto kritérium nebude úplně spolehlivé (např. Laird, 2004; Sackett, Borneman, & Connelly, 2008; Žoudlík, 2009). Sackett, Borneman a Connelly (2008) proto doporučují odhady souvislostí mezi výsledkem testů studijních předpokladů a školním výkonem srovnávat pouze v rámci souborů předmětů, které absolvovali všichni studenti společně, a u kterých by tedy jejich náročnost a hodnocení mělo být konzistentní. Tyto údaje ovšem pro naše analýzy k dispozici nemáme. Navíc je možné předpokládat, že na rozdíl od velkých zahraničních univerzit by počty studentů, kteří by absolvovali přesně tytéž předměty (i kdybychom se řídili jen orientačně, podle studovaných oborů) byly příliš nízké na to, aby bylo smysluplné počítat z nich regresní analýzy. Druhé hledisko omezení variability pak spočívá v tom, že korelace mezi známkovým průměrem a testem SPF jsou počítány pouze z dat uchazečů, kteří byli – buď zcela nebo z části – na základě těchto testů přijati (a nastoupili studium). Je tedy nasnadě, že jejich skóry budou obecně vyšší než celkové skóry všech (i nepřijatých) uchazečů. Tento druh omezení variability jsme se při výpočtech jednoduchých korelací pokusili zohlednit korekcí, hodnoty korelačních koeficientů se však změnily jen nepatrně, na $r = -0,10$ pro akad. rok 2011/2012 a $r = -0,26$ pro akad. rok 2012/2013. Stále se ještě nedají srovnávat s korigovanými korelacemi mezi SAT a FYGPA v hodnotě $r = -0,47$ (Sackett, Borneman, & Connelly, 2008). Jak ale poznamenává Říčan (1977), i za použití této korekce je dobré mít na paměti, že při nízké variabilitě prediktoru i kritéria případná nízká hodnota vzájemného korelačního koeficientu nemusí nutně znamenat, že test je

nevalidní, jako spíš že k jeho validizaci byly nedostatečně příznivé podmínky. Zjištěný korelační koeficient by se tak podobně jako u reliability dal považovat za odhad spodní hranice skutečné validity testu.

Pokud jde o predikční schopnosti testu SPF nad rámec dalších potenciálních prediktorů, zde byly nalezeny velmi rozdílné výsledky mezi ročníky SPF 2011 (akad. rok 2011/2012) a SPF 2012 (akad. rok 2012/2013).

Výsledek testu SPF v akademickém roce 2011/2012 nevysvětluje nad rámec průměrů známek u maturitní zkoušky ani nad rámec výsledků dalších součástí PZk, pakliže tyto byly povinné, prakticky žádnou varianci (ačkoliv přírůstek $R^2 = 0,008$ v souboru s maturitními známkami je významný statisticky). Naopak byl v tomto ročníku velmi výrazný efekt věku, který i po zadání všech ostatních proměnných dovedl objasnit 8-15% variance kritéria (tj. většinu z testovaného modelu), ačkoliv v celkovém souboru přijatých studentů věk se známkovým průměrem souvisel jen velmi slabě. Zdá se, že starší studenti dosáhli v tomto roce lepších známkových průměrů a věk byl dokonce i lepším prediktorem než-li maturitní průměr. Výjimkou z tohoto trendu byli studenti psychologie (soubor s váhou SPF 33%), u nichž přírůstek variance vysvětlený věkem již nebyl statisticky významný (ač i tak se jednalo o cca 4%) a naopak jejich maturitní průměr objasňoval 11% variance kritéria³⁶. Ten jinak konzistentně napříč ročníky vysvětloval asi 2% variance nad rámec ostatních proměnných a v kompletním modelu byl v akad. roce 2011/2012 lepším prediktorem než výsledek testu SPF a oborového testu (který taktéž nevysvětloval prakticky žádnou varianci).

Důvod, proč v tomto ročníku testy SPF ani oborové testy nevykazovaly téměř žádnou predikční schopnost, nám zůstává nejasný. Mohlo se jednat o efekt faktu, že šlo o první ročník používající vlastní testy SPF (navíc s vysokým podílem otázek ze všeobecného přehledu) a jejich tvůrci tedy ještě neměli odzkoušené vhodné typy položek. Také mohlo jít čistě o náhodný výkyv ve skutečných výkonech uchazečů. V neposlední řadě mohly být také výpočty prediktivní validity zatížené omezením variability, v tomto případě kritéria i prediktorů (zkoumané vzorky byly vždy výrazně menší než celkový počet uchazečů a nebylo vždy možné provést korekci pro omezení variability), ačkoliv při

³⁶ V budoucnu bychom se na studenty psychologie i z tohoto důvodu chtěli zaměřit blíže, neboť v akad. roce 2012/2013 již nebyli analyzováni samostatně – váha SPF byla posunuta na 30% a studenti psychologie tak byli zahrnuti do tohoto souhrnného vzorku. Tyto pozorované rozdíly by však mohly být odrazem vhodnosti doporučení Žoudlíka (2009) a Sacketta, Bornemana a Connollyho (2008) analyzovat obory samostatně.

splnění podmínek homoskedasticity a linearity by tento vliv měl být minimální (Eid, Gollwitzer, & Schmitt, 2010).

V akademickém roce 2012/2013 již výsledek testu SPF vysvětloval asi 5% variance kritéria nad rámec maturitního průměru a 15% variance nad rámec zbylé části přijímací zkoušky (dohromady 12% nad rámec maturitních známek + zbylé části PZk), přičemž byl současně silnějším prediktorem než absolvovaný oborový test/ústní zkouška a/nebo doložené certifikáty. Naopak vliv věku v tomto ročníku vymizel. Nabízíme tedy ke zvážení, zda i v oborech s povinnými dalšími částmi přijímací zkoušky nepřiložit testu SPF vyšší váhu.

Za zvážení stojí také zohlednění výsledků uchazečů u maturitní zkoušky. Jakkoliv podléhá známkování na střední škole (a u maturit tím spíše) stejnému potenciálnímu zkreslení jako známkování na VŠ, maturitní průměr poměrně konzistentně a nezávisle na ostatních prediktorech objasňoval 2% (či dokonce více) variance známek v prvním ročníku studia. V modelech bez oborového testu byl dokonce lepším prediktorem než body v testu SPF. Také Burton a Ramist uvádějí průměrné korelace mezi známkami ze střední školy a celkovým GPA v rozpětí $r = 0,34-0,57$, pro známky ze střední školy a průměr v prvním roce studia $r = 0,36$, což je vyšší souvislost než v případě korelace mezi testy SAT a (FY)GPA. Kombinací středoškolských známek a skóru SAT se korelace s GPA pohybuje kolem $r = 0,52$, korelace se známkami v prvním roce studia poté okolo $r = 0,44$. Kobrin, Patterson, Shaw, Mattern a Barbuti (2008) také shledali jako nejsilnější samostatný prediktor FYGPA známky ze střední školy (po korekci pro omezení variability $r = 0,54$). Nejlepší predikci FYGPA celkově zajistila kombinace známek ze SŠ a všech subtestů SAT (korigované $r = 0,62$). Stejně tak Richardson, Bond a Abraham (2012) uvádějí, že středoškolské známky jsou lepším prediktorem VŠ známek než testy SAT/ACT, ačkoliv vysvětlují rozdílný podíl variance (dohromady přibližně 22-25%). Podobně Štuka a Šimeček (2006) konstatují, že známky ze SŠ nesou poněkud odlišnou informaci než známky z přijímacích zkoušek, a zdají se být také lepšími samostatnými prediktory úspěšného dokončení studia. Sternberg (2004) dokonce udává, že při současném zadání známek ze střední školy a výsledků SAT jsou známky ze SŠ výrazně lepším prediktorem než SAT a dohromady vysvětlují 16,4% celkové variance. Tyto výsledky jsou velmi podobné našim nálezům a jejich možnou příčinou je určitá „obsahová podobnost“ obou proměnných, totiž že v obojím případě se jedná o známky – dosažené výkony ve zkouškových situacích. Můžeme uvažovat, do jaké míry tyto nejspíše odrážejí skutečné

dosavadní výkony (a nikoliv zkreslení způsobené známkujícím) uchazeče a jeho schopnosti, popřípadě motivaci (viz Camara, 2004a). V tomto ohledu je také zajímavé zjištění, že známkový průměr uchazečů přijatých bez PZk je nižší (lepší) než průměr uchazečů přijatých na základě přijímací zkoušky či odvolání. Jsme si vědomi, že rozdíly ve velikosti těchto srovnávaných skupin jsou opravdu vysoké a navíc se přijímání bez PZk patrně týká pouze vybraných oborů (jejichž celková či relativní náročnost oproti oborům jiným nám není známa), tuto souvislost by však bylo vhodné v budoucnosti lépe prozkoumat.

V této souvislosti však zůstává zásadní (a nezodpovězená) otázka, koho vlastně coby studenta na vysoké škole – či konkrétně FF UP – chceme. Pakliže to má být člověk, který bude získávat dobré známky, je možné diskutovat nejhodnější variaci na současnou podobu přijímacího řízení, totiž na váhy SPF, oborových testů, případně dalších osvědčení, certifikátů a dokladů o dosavadních výkonech (třeba i včetně známek z maturitních zkoušek). Pokud však zvolíme širší kritérium akademického úspěchu – např. v zahraničí doporučené dosažené úspěchy v oblasti vědy, umění či sportu, získání vysokých vedoucích pozic nebo společenskou angažovanost (např. College Entrance Examination Board, 2002; Schmitt, Oswald a Gillespie, 2004; aj.) – bude potřeba zauvažovat o *nekognitivních* prediktorech. To s sebou samozřejmě ponese další obtíže, mimo jiné například vyřešení operacionalizace těchto proměnných a jejich spolehlivého měření. Sebeuposuzovací inventáře a škály pro tyto oblasti sice již existují, trpí však citlivostí vůči zkreslení ve směru sociální žádoucnosti (např. Camara, 2004b; Sternberg, 2004). Za zvážení by také stálo ověřování tvořivosti (viz Sternberg, 2004) či osobnostních vlastností uchazečů, které však může narazit na nelibost veřejnosti a nařčení ze subjektivity a neférovosti. V první řadě však bude potřeba vůbec definovat cíle a hodnoty Univerzity Palackého a Filozofické fakulty, a z nich následně odvozovat požadavky na uchazeče, které hledá. Zároveň bychom jako tvůrci těchto požadavků měli mít na mysli, za jakým účelem chceme, aby měl uchazeč 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? S ohledem na tato kritéria pak také bude možné přesněji formulovat a ověřovat žádoucí psychometrické vlastnosti metod, které k jejich zjišťování budou použity.

11 Závěry

Z výsledků této práce vyplývá, že zjištěná reliabilita testů SPF v uplynulých ročnících se z pohledu vnitřní konzistence pohybovala okolo spodní stanovené hranice, od níž je možné testy považovat za reliabilní/vnitřně konzistentní. Odhady paralelity párových (lichých a sudých) verzí testu pomocí srovnání jejich koeficientů vnitřní konzistence, průměrů a směrodatných odchylek poukazují na fakt, že při liberálnější hladině spolehlivosti alfa by mezi některými obsahově identickými verzemi mohl být shledán rozdíl. Vzhledem k logistice průběhu administrace testů SPF však není jasné, zda tato odlišnost spočívá v testech samotných či spíše v uchazečích, kteří dané verze absolvují.

Také posouzení obsahové validity se ukázalo jako obtížné. Nalezené faktory, pracovně nazvané „logicko-analytické myšlení a představivost“ a „sečtělost“, vysvětlují jen velmi nízký podíl celkové variance a také jednotlivé položky málokdy dosahují hodnot faktorových zátěží potřebných pro jednoznačné konstatování, že tyto položky jsou nalezenými faktory syceny. Zůstává otázkou, zda je tento fakt následkem nízké variability pravých skóre uchazečů nebo nízkou homogenitou testů samotných.

Pomocí položkové analýzy se podařilo identifikovat některé typy položek, které konzistentně dosahují střední obtížnosti a zároveň dobré rozlišovací schopnosti (např. rotace komplexnějších 2D objektů v části prostorové představivosti, slovní úlohy v části logicko-analytické). Zároveň byly odhaleny typy položek, jejichž přítomnost v testu je spíše „doplňková“ (např. doplňování řad čísel a písmen) a v budoucích verzích by bylo možné nahradit je položkami kvalitnějšími (např. doplňováním čísel do obrazců).

V analýzách prediktivní validity testů SPF vůči známkám v prvním ročníku VŠ a také inkrementální validity těchto testů oproti zbylé části přijímací zkoušky, maturitnímu průměru, popř. věku, se ukázal výrazný rozdíl mezi akademickým rokem 2011/2012 a akad. rokem 2012/2013. V prvním jmenovaném testy neobjasňovaly prakticky žádnou varianci kritéria a naopak byl velmi výrazný vliv věku. V akademickém roce 2012/2013 již výsledek testu SPF vysvětloval signifikantní podíl variance kritéria (i nad rámec ostatních proměnných), přičemž byl současně silnějším prediktorem než absolvovaný oborový test/ústní zkouška a/nebo doložené certifikáty. Naopak vliv věku v tomto ročníku vymizel. Současně byl konzistentním prediktorem známek v prvním roce studia maturitní průměr, a to velmi výrazně zvláště u studentů psychologie v akad. roce 2011/2012. Doporučujeme

proto zvážit vyšší zohlednění testu SPF, popř. i maturitních průměrů, v přijímacím řízení na FF UP.

Naše zjištění mají samozřejmě určitá interpretační omezení, způsobená především některými statistickými fenomény (omezení variability), metodologickými problémy (nepřístupnost některých dat) a v neposlední řadě problematickou definicí samotného konstruktů studijních předpokladů. I z tohoto důvodu doporučujeme opatrnost při interpretaci především výsledků faktorové analýzy a srovnání párových verzí testů. Zároveň považujeme za vhodné v budoucnu sledovat průběh i dalších ročníků přijímacích zkoušek a testů SPF, popř. zavést další kritéria akademického úspěchu, a pokusit se o podrobnější rozbor z hlediska jednotlivých oborů na FF UP. Prvním krokem by mohly být např. analýzy na souboru studentů psychologie. Celkově tedy považujeme výsledky této práce za první předběžné odhady psychometrických vlastností Testů předpokladů ke studiu na FF UP (SPF).

Souhrn

Cílem této práce bylo ověření psychometrických vlastností (reliability a validity) Testů předpokladů ke studiu na FF UP (SPF), které vytvářejí odborníci z řad zaměstnanců FF UP a které jsou povinnou součástí přijímacího řízení na bakalářské obory Filozofické fakulty Univerzity Palackého od akad. roku 2011/2012. Obsahují 50 otázek, na jejichž vyplnění má uchazeč 60 minut, přičemž z pěti odpovědí je vždy pouze jedna správná. V uplynulých letech bylo administrováno vždy 12-16 verzí testu, přičemž pokaždé byly dvě verze párové (obsahovaly stejné otázky, pouze v jiném pořadí a/nebo s odlišným označením správné odpovědi).

Důvodem této analýzy byla snaha o zjištění, zda je nově vyvinutý test SPF kvalitním nástrojem výběru uchazečů, který umožňuje předpovídat jejich budoucí úspěch ve studiu. V teoretické části této práce zároveň uvádíme, že samotná definice „akademického úspěchu“ je problematickým a komplexním tématem. V zahraničí se v této souvislosti diskutují kognitivní (výkonové, např. znalostní) a nekognitivní (osobnostní a motivační) faktory a ukazatele „úspěšného studia“ (např. College Entrance Examination Board, 2002). Současně s tím probíhá diskuze o nejvhodnějších prediktorech akademického úspěchu. Velká pozornost je v ní věnována konstruktů studijních předpokladů, ačkoliv ten sám doposud není konkrétně definovaný a jeho odlišnost od konceptu obecné inteligence (g-faktoru) zůstává nejasná (Gottfredsonová, 2004). Většina testů studijních předpokladů se tak skládá z více či méně si navzájem podobných oddílů zaměřených na specifické schopnosti - řešení problémů, logické, analytické, kritické či abstraktní myšlení, prostorovou představivost apod. Kritériem jejich validity (a zároveň operacionalizací akademického úspěchu) je poté nejčastěji fakt, jak dobře dovedou předpovídat známky studentů na VŠ, případně konkrétně v prvním ročníku (Camara, 2004a).

Výsledky validizačních studií testů studijních předpokladů (v zahraničí např. test SAT, u nás např. testy OSP společnosti Scio) se různí. Nalezené korelace mezi výsledky testu a známkami v prvním roce studia sahají od hodnot $r = 0,13$ (Štuka, 2012) po hodnoty $r > 0,45$ (Burton, & Ramist, 2001; Sackett, Borneman, & Connelly, 2008 aj.). Někteří autoři navíc upozorňují na lepší predikční schopnosti oborových testů a známek ze střední školy než testů studijních předpokladů (např. Höschl, & Kožený, 1997; Konečný, Basl, & Mysliveček, 2010; Zvára, & Anděl, 2001).

Rozhodli jsme se proto ověřit psychometrické vlastnosti testů SPF jakožto testů nových, na jejichž výsledcích závisí budoucnost několika tisícovek uchazečů. Vycházeli jsme přitom z postupů v odborné literatuře (např. Crocker, & Algina, 2008; Urbánek, Denglerová, & Širůček, 2011) doporučených pro ověřování reliability a validity psychodiagnostických nástrojů, zvláště pak výkonových testů. Reliabilitu ve smyslu vnitřní konzistence jsme hodnotili pomocí Kuder-Richardsonova vzorce 20. Paralelitu jednotlivých verzí testu nebylo možné posuzovat běžným způsobem (administrací více verzí též uchazečům), rozhodli jsme se tedy pro odhad srovnáním průměrů (dvouvýběrový t-test), směrodatných odchylek (F-test pro rozptyl) a koeficientů vnitřní konzistence (z-test) párových verzí testu. K posouzení obsahové validity testů jsme zvolili explorační faktorovou analýzu za použití metody hlavních os a rotace typu Promax, a dále položkovou analýzu na principech klasické testové teorie ke zjištění obtížnosti a rozlišovací schopnosti položek. Prediktivní validita byla posuzována vůči známkovému průměru v prvním roce studia, a to pomocí jednoduchých Pearsonových korelací a prostřednictvím tří modelů krokové regresní analýzy, v nichž byly jako prediktory zvoleny kromě výsledků v testu SPF ještě výsledek ve zbylé části přijímací zkoušky (pakliže tuto uchazeč absolvoval), maturitní průměr a věk.

Pracovali jsme s dvěma typy datových souborů, jednak s maticemi odpovědí všech účastníků testů SPF v uplynulých ročnících (SPF 2011: 4.480 účastníků; SPF 2012: 4.459 účastníků; SPF 2013: 4.231 účastníků), jednak s daty přijatých a zapsaných uchazečů z let 2011 a 2012, pro něž byl k dispozici známkový průměr v prvním roce studia ($n_{2011} = 1.250$; $n_{2012} = 1.331$). Tento soubor byl dále rozdělen podle dostupných údajů o dalších proměnných (váha testu SPF v celkové přijímací zkoušce, maturitní průměr studentů) a takto vzniklé soubory byly samostatně analyzovány.

Z výsledků odhadů reliability vyplývá, že vnitřní konzistence testů SPF se v uplynulých ročnících pohybovala okolo spodní hranice, od níž je možné testy považovat za spolehlivé. Odhady paralelity párových (lichých a sudých) verzí testu poukazují na fakt, že při liberálnější hladině spolehlivosti alfa by mezi některými obsahově identickými verzemi mohl být shledán rozdíl. Vzhledem k logistice průběhu administrace testů SPF však není jasné, zda tato odlišnost spočívá v testech samotných či spíše v uchazečích, kteří dané verze absolvují.

Z hlediska obsahové validity také nalezené faktory, pracovně nazvané „logicko-analytické myšlení a představivost“ a „sečtělost“, vysvětlují jen velmi nízký podíl celkové variance skóru. Jednotlivé položky málokdy dosahují hodnot faktorových zátěží potřebných pro jednoznačné konstatování, že tyto položky jsou nalezenými faktory syceny. Za pomoci položkové analýzy se však podařilo identifikovat alespoň některé typy položek, které konzistentně dosahují střední obtížnosti a zároveň dobré rozlišovací schopnosti. Zároveň byly odhaleny typy položek, jejichž přítomnost v testu je spíše nadbytečná a v budoucích verzích by bylo možné nahradit je položkami kvalitnějšími.

V analýzách prediktivní validity testů SPF vůči známám v prvním ročníku VŠ se ukázal výrazný rozdíl mezi akad. rokem 2011/2012 a akad. rokem 2012/2013. V prvním jmenovaném testu neobjasňovaly prakticky žádnou varianci kritéria a naopak byl velmi výrazný vliv věku. V akademickém roce 2012/2013 již výsledek testu SPF vysvětloval signifikantní podíl variance kritéria (i nad rámec ostatních proměnných), přičemž byl současně silnějším prediktorem než zbylá část přijímací zkoušky. Současně byl konzistentním prediktorem známek v prvním roce studia maturitní průměr, a to velmi výrazně zvláště u studentů psychologie v akad. roce 2011/2012. Doporučujeme proto zvážit vyšší zohlednění testu SPF, popř. i maturitních průměrů, v přijímacím řízení na FF UP.

Naše zjištění mají interpretační omezení daná především statistickými fenomény (omezení variability), metodologickými problémy (nepřístupnost některých dat) a v neposlední řadě problematickou definicí samotného konstruktů studijních předpokladů. I z tohoto důvodu doporučujeme opatrnost při interpretaci především výsledků faktorové analýzy a srovnání párových verzí testů. Zároveň považujeme za vhodné v budoucnu sledovat průběh i dalších ročníků přijímacích zkoušek a testů SPF, popř. zavést další kritéria akademického úspěchu, a pokusit se o podrobnější rozbor z hlediska jednotlivých oborů na FF UP.

Seznam použitých zdrojů a literatury

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/.

Bollinger, L.C. (2004). Competition in higher education and admissions testing. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (3-12). New York: Routledge.

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>.

Camara, W.J. (2004a). Broadening criteria of college success and the impact of cognitive predictors. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (53-79). New York: Routledge.

Camara, W.J. (2004b). Broadening predictors of college success. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (81-105). New York: Routledge.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge: Cambridge University Press.

College Entrance Examination Board. (2002). *Best practices in admissions decisions: A report on the third College board conference on admission models*. Získáno z <http://research.collegeboard.org/sites/default/files/publications/2012/7/misc2002-1-best-practices-admissions-decisions.pdf>.

Conley, D. T. (2003). *Understanding university success: a report from Standards for Success: a project of the Association of American Universities and the Pew Charitable Trusts*. Eugene, OR: Center for Educational Policy Research.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage.

Eid, M., Gollwitzer, M., & Schmitt, M. (2010). *Statistik und Forschungsmethoden Lehrbuch; mit Online-Materialien* (1. Aufl. ed.). Weinheim: Beltz.

Filozofická fakulta Univerzity Palackého v Olomouci. (2013). *O fakultě*. Získáno z <http://www.ff.upol.cz/menu/o-fakulte/>.

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>.

Geiser, S. & Studley, R. (2001). *UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California*. Oakland: University of California, Office of the President.

Gorin, J. S. (2007). Reconsidering Issues in Validity Theory. *Educational Researcher*, 36(8), 456-462.

Gottfredson, L. S. (2004). Schools and the g factor. *The Wilson Quarterly*, Summer, 35-45.

Grigárková, P. (4. června 2012). *Charakteristika hodnocení testů OSP a OT. Hodnocení TSP a OT pro jednoobory a dvouobory z přijímacího řízení pro akademický rok 2012/2013*. Získáno z <http://www.ff.upol.cz/skupiny/zajemcum-o-studium/bakalarske-a-magisterske-studium/testy-prijimaciho-rizeni/>.

Grigárková, P. (18. října 2013). *Zpráva o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2013/2014*. Získáno z <http://www.ff.upol.cz/menu/prijimaci-rizeni/bakalarske-a-magisterske-studium/#c24668>.

Hartl, P. (2004). *Stručný psychologický slovník*. Praha: Portál.

Höschl, C., & Kožený, J. (1997). Predicting academic performance of medical students: the first three years. *American Journal of Psychiatry*, 154, 87–92.

Howitt, D., & Cramer, D. (2005). *Introduction to research methods in psychology*. Harlow, England: Pearson/Prentice Hall.

Charvát, M., Opletalová, V., Tošenovská, M., Viktorová, L., & Vobořil, L. (2014). *Tvorba a administrace testů studijních předpokladů*. Rukopis v přípravě. Olomouc: VUP.

Jelínek, M., Květoň, P., & Vobořil, D. (2011). Testování v psychologii: Teorie odpovědi na položku a počítačové adaptivní testování. Praha: Grada.

Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association. Získáno 4. dubna 2013 z <https://law.resource.org/pub/us/cfr/ibr/001/aera.standards.1999.pdf>.

Konečný, T., Basl, J., Mysliveček, J. (2010). Přejít mezi střední a vysokou školou a role přijímacího řízení. *Sociologický časopis*, 46, 43-72.

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>.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All?. *Journal of Personality and Social Psychology*, 86 (1),148-161. Získáno z <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.317.9553&rep=rep1&type=pdf>.

Laird, R. (2004). What is it we think we are trying to fix and how should we fix it? A view from the admissions office. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (13-32). New York: Routledge.

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>.

Masarykova univerzita v Brně. (20. května 2013). *Test studijních předpokladů*. Získáno z <http://www.muni.cz/tsp>.

Matějů, P., Ježek, F., Münich, D., Polechová, P., Slovák, J., Straková, J.,... Zrzavý, J. (2009). *Bílá kniha terciárního vzdělávání*. Praha: MŠMT.

National Association for College Admission Counseling. (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.

Opletalová, V., & Pavelková, Z. (2013). *Test předpokladů ke studiu s důrazem na orientaci v humanitních a sociálně-vědných disciplínách (SPF)*. Nepublikovaný rukopis, Katedra germanistiky, Univerzita Palackého v Olomouci.

Pavličková, H. (nedat.). *Několik postřehů k osobnosti vysokoškolského studenta*. Získáno 27. března 2013 z <http://www.agris.cz/clanek/142083>.

Prudký, L., Pabian, P., & Šima, K. (2010). *České vysoké školství: Na cestě od elitního k univerzálnímu vzdělávání 1989-2009*. Praha: Grada.

Přírodovědecká fakulta UK v Praze. (nedat.) *Přijímací řízení do bakalářských studijních programů*. Získáno 10. února 2014 z <http://www.natur.cuni.cz/fakulta/uchazeci/bakalarske-studium/prijimaci-rizeni/bakalarske-studium-2010-2011/#zkouska>.

Richardson, M., Bond, R., & Abraham, C. (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin*, 138(2), 353-387.

Rychlík, M. (29. ledna 2014). *Historik Miller: Univerzity by měly usilovat o lepší svět. Česká pozice*. Získáno z <http://www.ceskapozice.cz/domov/veda-vzdelavani/historik-miller-univerzity-mely-usilovat-o-lepsi-svet>.

Říčan, P. (1977). *Úvod do psychometrie*. Bratislava: Psychodiagnostické a didaktické testy.

Sackett, P. R., Borneman, M. J., & Connelly, B.S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, 63(4), 215-227.

Scio. (květen 2012). *Srovnávací analýza OSP a IQ*. Získáno z: <http://www.scio.cz/vyzkum/analyzy/osp-iq.asp>.

Scio. (15. ledna 2013). *Souběžná validita testů SAT a OSP*. Získáno z https://www.scio.cz/download/SAT_vs_OSP_soubezna_validita.pdf.

Scio. (20. května 2013). *Základní klasifikace testů*. Získáno z: http://www.scio.cz/vyzkum/tvorba_testu/teorie_testu/typy_testu.asp.

Schmidt, F. L., Oh, I., & Le, H. (2006). Increasing The Accuracy Of Corrections For Range Restriction: Implications For Selection Procedure Validities And Other Research Results. *Personnel Psychology*, 59(2), 281-305.

Schmitt, N., Oswald, F.L., & Gillespie, M.A. (2004). Broadening the Performance Domain in the Prediction of Academic Success. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (3-12). New York: Routledge.

Schultz, K.S., & Whitney, D.J. (2005). *Measurement theory in action: Case studies and exercises*. Thousand Oakes: Sage Publications.

Sternberg, R. J. (2004). *Augmenting the SAT Trough Assessments of Analytical, Practical and Creative Skills*. In W.J. Camara & E.W. Kimmel (Eds.), *Choosing students: higher education admissions tools for the 21st century* (159-176). New York: Routledge.

Škaloudová, A. (2010). *Explorativní faktorová analýza: Základní pojmy*. Získáno 2. února 2014 z http://userweb.pdf.cuni.cz/kpsp/skalouda/fa/zakl_pojmy.htm.

Štuka, Č. (2012). *Úspěšnost studia z pohledu moderních metod analýzy dat*. (Nepublikovaná disertační práce). Univerzita Karlova v Praze.

Štuka, Č., & Šimeček, P. (2006). Studium souvislosti mezi úspěšností studia medicíny, známkami studentů na střední škole a výsledky přijímacích zkoušek. *Sborník Medsoft*, získáno 10.února 2014 z <http://atrey.karlin.mff.cuni.cz/~simecek/skola/StukaSimecek.doc>.

Tošenovská, M. (2014). *Profil uchazeče o studium na FF UP*. (Nepublikovaná bakalářská práce). Univerzita Palackého v Olomouci.

Univerzita Palackého v Olomouci. (2013). *O univerzitě*. Získáno z <http://www.upol.cz/o-univerzite/>.

Urbánek, T., Denglerová, D., Širůček, J. (2011). *Psychometrika: měření v psychologii*. Praha: Portál.

Urbášek, P. (2008) *Vysokoškolský vzdělávací systém v letech tzv. normalizace*. Olomouc: Univerzita Palackého v Olomouci.

Vyhláška č. 343/2002 Sb. o postupu a podmínkách při zveřejnění průběhu přijímacího řízení na vysokých školách.

WikiSkripta, projekt sítě lékařských fakult MEFANET. (20.dubna 2013). *Testování při výuce medicíny: Konstrukce a analýza testů na lékařských fakultách*. Získáno z: <http://www.wikiskripta.eu/index.php/Fórum:Testy>.

Willingham, D. (18. února 2013). *What predicts college GPA?* Získáno 2. února 2014 z <http://www.danielwillingham.com/1/post/2013/02/what-predicts-college-gpa.html>.

Zákon č. 111/1998 Sb. o vysokých školách.

Zákon č. 2/1993 Sb. - Listina základních práv a svobod.

Zpráva o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2011/2012. (nedat.). Získáno 10. února 2014 z www3.upol.cz/fileadmin/user_upload/.../FF-zprava-prij2011-2012.doc.

Zpráva o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2012/2013.(nedat.). Získáno 10. února 2014 z http://www.ff.upol.cz/fileadmin/user_upload/FF-dokumenty/prijimacky-zpravy/2012/2012-Zprava_o_prubehu_prij.doc.

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>.

Žoudlík, J. (2009). *Přijímací zkoušky na vysokou školu jako prediktor akademické úspěšnosti*. (Nepublikovaná magisterská práce). Masarykova Univerzita v Brně.

Abstrakt diplomové práce

Název práce: Reliabilita a validita SPF (Test předpokladů ke studiu na FF UP)

Autor práce: Lucie Viktorová

Vedoucí práce: Mgr. Miroslav Charvát, Ph.D.

Počet stran a znaků: 109 stran, 233.927 znaků (vč. mezer)

Počet příloh: 10

Počet titulů použité literatury: 63

Abstrakt (800–1200 zn.): Cílem této práce bylo ověření psychometrických vlastností (reliability a validity) Testů předpokladů ke studiu na FF UP (SPF), které jsou povinnou součástí přijímacího řízení na bakalářské obory Filozofické fakulty Univerzity Palackého v Olomouci od akad. roku 2011/2012. K odhadu reliability byl využit vzorec KR-20 a srovnání průměrů, směrodatných odchylek a koeficientů vnitřní konzistence párových verzí testu. K posouzení obsahové validity testů jsme zvolili explorační faktorovou analýzu a dále položkovou analýzu na principech klasické testové teorie. Prediktivní validita vůči známkovému průměru v prvním roce studia byla posouzena pomocí krokové regresní analýzy, v nichž byly jako prediktory zvoleny kromě výsledků v testu SPF ještě výsledek ve zbylé části přijímací zkoušky (pakliže tuto uchazeč absolvoval), maturitní průměr a věk. Z výsledků práce vyplývá, že vnitřní konzistence a obsahová validita testů SPF má ještě prostor k růstu. Z hlediska prediktivní validity lépe dopadl test SPF 2012, který byl zároveň silnějším prediktorem než zbylá část přijímací zkoušky. Doporučujeme proto zvážit vyšší zohlednění testu SPF, popř. i maturitních průměrů, v přijímacím řízení na FF UP.

Klíčová slova: studijní předpoklady, reliabilita, validita

Abstract of thesis

Title: Reliability and validity of the SPF (Learning potential test for FF UP)

Author: Lucie Viktorová

Supervisor: Mgr. Miroslav Charvát, Ph.D.

Number of pages and characters: 109 pages, 233.927 characters

Number of appendices: 10

Number of references: 63

Abstract (800–1200 characters): This thesis aims to test the psychometric characteristics – reliability and validity – of the Learning potential test for FF UP (SPF). This is a mandatory part of the selection process for bachelor studies at the Philosophical faculty of the Palacký University in Olomouc since the year 2011/2012. To estimate the test reliability, we used the KR-20 formula and comparisons of the means, standard deviations and internal reliability coefficients of the corresponding test versions. For content validity estimation we used exploratory factor analysis and item analysis on the base of classical test theory. The ability to predict students' first year grade point average was estimated with blockwise regression with the predictors: SPF score, score from other parts of the entrance examination (if present), grades from the high school-leaving certificate, and age. The results show that the internal consistency and content validity of the SPF still has a potential to grow. As for predictive validity, SPF 2012 was a better predictor than the rest of the entrance examination that year. We recommend considering giving the SPF a higher value in the entrance examination for FF UP and possibly considering high-school leaving certificate grades as well.

Key words: learning potential/aptitude, reliability, validity

Přílohy diplomové práce

Příloha č. 1: Zadání diplomové práce

Univerzita Palackého v Olomouci
Filozofická fakulta
Akademický rok: 2012/2013

Studijní program: Psychologie
Forma: Prezenční
Obor/komb.: Psychologie (PS)

Podklad pro zadání DIPLOMOVÉ práce studenta

PŘEDKLÁDÁ	ADRESA	OSOBNÍ ČÍSLO
VIKTOROVÁ Lucie	U Klimentky 3, Praha - Smíchov	F08155

TÉMA ČESKY:

Reliabilita a validita SPF (Test předpokladů ke studiu na FF UP)

NÁZEV ANGLICKY:

Reliability and validity of the SPF (Learning potential test for FF UP)

VEDOUcí PRÁCE:

Mgr. Miroslav Charvát, Ph.D. - PCH

ZÁSADY PRO VYPRACOVÁNÍ:

Testy studijních předpokladů jsou v různé podobě součástí přijímacího řízení na vysoké školy v ČR i zahraničí již řadu let. Jejich cílem je ověřit verbální, matematické a kritické myšlení uchazečů coby nezbytné schopnosti pro studium na VŠ. Počínaje akademickým rokem 2011/2012 zavedla i Univerzita Palackého jako povinnou součást přijímacího řízení společnou pro všechny obory Filozofické fakulty vlastní Test předpokladů ke studiu na FF UP (SPF). Vzhledem k tomu, že se jedná o relativně nový a do velké míry psychologicky orientovaný test, je oprávněná otázka prověření jeho dosavadní reliability a validity. Za tímto účelem vzniká i současná diplomová práce, v níž se pokusím nastínit koncepty reliability a validity v psychometrii, stejně jako některé možnosti jejich statistického ověřování, a aplikovat je na data získaná od účastníků SPF v předchozích letech.

Cílem výzkumu je tedy ověřit reliabilitu a validitu proběhlých testů SPF na FF UP, a to na datovém souboru získaném od cca 4900 uchazečů o studium v každém z uplynulých ročníků. Data budou analyzována mj. pomocí faktorové, regresní a položkové analýzy. Očekávaným

výstupem bude vedle ověření psychometrických vlastností výše zmíněných testů např. identifikace problematických a naopak dobře diferencujících testových položek a prozkoumání predikční schopnosti SPF vzhledem k průměrnému prospěchu přijatých uchazečů v prvním roce studia. Tyto údaje by měly napomoci ke zkvalitnění tvorby přijímacích testů na FF UP a v budoucnu např. k cílenému testování žádaných profilů studentů jednotlivých oborů FF UP. Svůj postup práce budu pravidelně konzultovat s vedoucím práce a diplomovou práci sepiší se všemi náležitostmi podle norem katedry. V případě kvalitního zpracování budu svá zjištění publikovat v odborném tisku.

SEZNAM DOPORUČENÉ LITERATURY:

Cronbach, L. J., Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. Guilford, J. P. (1948). Factor analysis in a test-development program. *Psychological Review*, 55, 79-94.

Höschl, C., Kožený, J. (1997). Predicting academic performance of medical students: the first three years. *American Journal of Psychiatry*, 154, 87-92.

Konečný, T., Bašl, J., Mysliveček, J. (2010). Přejít mezi střední a vysokou školou a role přijímacího řízení. *Sociologický časopis*, 46, 43-72.

Říčan, P. (1977). Úvod do psychometrie. Bratislava: Psychodiagnostické a didaktické testy.

Zvára, K., Anděl, J. (2001). Souvislost výsledků přijímacího řízení s úspěšností studia na MFF. *Pokroky matematiky, fyziky a astronomie*, 46 (4), 304-312.

Podpis studenta: **Datum:**

Podpis vedoucího práce: **Datum:**

Příloha č. 2: Přehled Bc. oborů a užitých způsobů přijímacích zkoušek

Tab. I: Bakalářské obory nabízené FF UP v akad. letech 2011/2012-2013/2014 a způsob přijímacích zkoušek

Bakalářské studium	Akad. rok 2011/2012	Akad. rok 2012/2013	Akad. rok 2013/2014
Prezenční jednooborové:			
Anglická filologie	SPF, OT	SPF, OT, CER	SPF, OT
Angličtina se zaměřením na aplikovanou ekonomii	SPF, OT, CER	X	X
Angličtina se zaměřením na tlumočení a překlad	SPF, OT, UZ	SPF, OT	SPF, OT, UZ
Archivnictví	SPF, OT	SPF, OT	SPF, OT
Česká filologie se zaměřením na editorskou činnost	SPF	SPF, OT	SPF
Česká filologie	SPF	SPF, OT	SPF
Čínská filologie	X	SPF, OT	SPF
Dějiny výtvarných umění	SPF, UZ	SPF, OT	SPF, UZ
Dutch Language, Literature and Culture in a Central European Context	SPF	SPF, OT	X
Francouzština se zaměřením na aplikovanou ekonomii	SPF, CER, MAT	SPF, OT	SPF, OT
Historie	SPF, OT	SPF, OT	SPF, OT
Japonská filologie	X	SPF, OT	SPF, OT
Kulturní antropologie	SPF	SPF	SPF
Muzikologie	SPF	SPF	SPF
Německá filologie	SPF, OT	SPF, OT	SPF, OT, CER, MAT, VYS
Nizozemská filologie	X	X	SPF
Nizozemština pro hospodářskou praxi	SPF	SPF	X
Politologie a evropská studia	SPF, OT	SPF, OT	SPF, OT
Polská filologie v kontextu evropské kultury a literatury	SPF	X	SPF
Polština se zaměřením na hospodářsko-právní a turistickou oblast	SPF	SPF	SPF
Psychologie	SPF, UZ	SPF, UZ	SPF, UZ
Ruská filologie v kontextu evropské kultury a literatury	SPF, OT	SPF, OT	SPF, OT
Ruština se zaměřením na hospodářsko-právní a turistickou oblast	SPF, OT	SPF, OT	SPF, OT
Španělská filologie	SPF, CER, MAT	SPF	SPF, OT
Teorie a dějiny dramatických umění	SPF, UZ	SPF, UZ	SPF, UZ
Ukrajinaština se zaměřením na hospodářsko-právní a turistickou oblast	SPF	SPF	SPF

Uměnovědná studia	SPF	SPF	SPF
Žurnalistika	SPF	SPF	SPF
Prezenční dvouoborové:			
Andragogika	X	SPF	SPF
Andragogika a Sociologie (jen tato komb.)	SPF	X	X
Anglická filologie	SPF, OT	SPF, OT	SPF, OT, CER
Aplikovaná ekonomická studia	SPF	SPF	SPF
Archeologie	SPF, OT	SPF, OT	SPF, OT
Archivnictví	X	SPF, OT	SPF, OT
Česká filologie	SPF	SPF	SPF
Čínská filologie	SPF	SPF	SPF
Divadelní věda	SPF, UZ	SPF, UP	SPF, UZ
Filmová věda	SPF, UZ	SPF, UP	SPF, UZ
Filozofie	SPF, UZ	SPF, UP	SPF, UZ
Francouzská filologie	SPF, CER, MAT	SPF	SPF, OT
Historie	SPF	SPF	SPF
Italská filologie	SPF, CER, MAT	SPF	SPF
Japonská filologie	SPF	SPF	SPF
Japonština pro hospodářskou praxi	SPF	SPF	SPF
Judaistika: Židovská a izraelská studia	X	X	SPF
Kulturní antropologie	SPF	SPF	SPF
Latinská filologie	SPF, CER, MAT	SPF	SPF
Muzikologie	SPF	SPF	SPF
Německá filologie	SPF, OT, CER, MAT	SPF, OT, CER, MAT	SPF, OT, CER, MAT, VYS
Nizozemská filologie	SPF	SPF	SPF
Obecná lingvistika a teorie komunikace	X	X	SPF
Politologie a evropská studia	X	X	SPF, OT
Polská filologie	SPF	SPF	SPF
Portugalská filologie	SPF	SPF	SPF
Ruská filologie	SPF, OT	SPF, OT	SPF, OT
Sociologie	X	SPF	SPF
Španělská filologie	SPF, CER, MAT	SPF	SPF, OT
Ukrajinská filologie	SPF	SPF	SPF
Uměnovědná studia	SPF	SPF	SPF
Žurnalistika	SPF	SPF	SPF
Kombinované jednooborové			
Andragogika v profilaci na personální management	SPF, OT	SPF, OT	SPF, UZ

Muzikologie	SPF	SPF	SPF
Psychologie	SPF, UZ	SPF, UZ	SPF, UZ
Sociální práce	SPF, OT	SPF, OT	SPF, OT
Kombinované dvouoborové			
Divadelní věda – Filmová věda	SPF, UZ	SPF, UZ	SPF, UZ

Pozn.: CER = certifikát z jazyka; MAT = maturitní vysvědčení z předchozích let studia; OT= oborový test; SPF= test předpokladů ke studiu na FF UP; UZ = ústní zkouška; VYS = vysvědčení z posledního ročníku SŠ; X= obor nebyl v uvedeném roce nabízen. Tabulka je založena na analýzách Grigárkové (18. října 2013), Zprávě o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2011/2012 (nedat.) a Zprávě o průběhu přijímacího řízení na Filozofické fakultě UP v Olomouci pro akademický rok 2012/2013 (nedat.).

Příloha č. 3: Ukázka datových matic

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	
Identifikátor	Verze	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50		
352	477226	1	1	1	0	1	1	1	0	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	0	0	1	0	0				
353	477542	1	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0		
354	477654	1	1	1	0	0	0	1	1	1	0	0	1	0	0	1	0	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	0	0	1	0		
355	478290	1	1	1	1	0	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0			
356	478350	1	1	1	0	0	1	0	1	1	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0		
357	478466	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0			
358	478818	1	0	1	1	0	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	0	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0		
359	479070	1	0	1	0	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	0	0	0	1	0	0			
360	479210	1	0	1	0	0	1	1	1	1	1	1	0	1	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0		
361	479774	1	1	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1		
362	479846	1	0	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1	1	1	0	0	1	0	0	
363	479938	1	0	0	1	1	0	1	0	1	1	0	0	1	1	0	0	0	1	0	1	0	0	1	1	1	1	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0	0	
364	480984	1	0	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0		
365	481208	1	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0		
366	481676	1	1	1	0	0	0	1	1	1	0	1	0	0	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0		
367	482546	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	
368	482558	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	1	1	1	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0	
369	56282	2	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	0	1	0	0	1	0	0	1	1	0	0	1	1	0	0	0	1	0	0	0	1	0	0	
370	143720	2	1	0	1	0	1	1	1	1	1	0	0	0	1	1	1	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	1	0	0	1	1	0	1	1	0	
371	147924	2	1	0	0	1	1	0	1	0	1	1	0	0	1	1	1	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	0	0	
372	148680	2	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	0	
373	232098	2	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	
374	250762	2	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0	1	0	0	1	1	1	0	1	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	0	0	
375	256648	2	0	1	0	1	0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0	1	0	1	1	0	1	0	
376	258476	2	0	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	1	0	1	1	1	0	0	1
377	264838	2	0	1	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Obr. I: Ukázka datové matice testu SPF 2013

Příloha č. 4: Přehled Bc. oborů s nesrovnalostmi ve vahách částí PZk

Tab. II: Seznam oborů s jiným než deklarovaným poměrem SPF:OT/ÚZ a/nebo max. celkovým počtem bodů

Obor	Akad. rok
Angličtina se zaměřením na komunitní tlumočení a překlad	2012/2013
Aplikovaná ekonomická studia - Filmová věda	2011/2012
Aplikovaná ekonomická studia - Filmová věda	2012/2013
Anglická filologie - Aplikovaná ekonomická studia	2011/2012
Anglická filologie - Aplikovaná ekonomická studia	2012/2013
Anglická filologie - Archeologie	2012/2013
Anglická filologie - Čínská filologie	2011/2012
Anglická filologie - Čínská filologie	2012/2013
Anglická filologie - Francouzská filologie	2011/2012
Anglická filologie - Francouzská filologie	2012/2013
Anglická filologie - Filmová věda	2011/2012
Anglická filologie - Filmová věda	2012/2013
Anglická filologie - Historie	2012/2013
Anglická filologie - Italská filologie	2012/2013
Anglická filologie - Japonská filologie	2011/2012
Anglická filologie - Japonská filologie	2012/2013
Anglická filologie - Japonština pro hospodářskou praxi	2011/2012
Anglická filologie - Japonština pro hospodářskou praxi	2012/2013
Anglická filologie - Latinská filologie	2011/2012
Anglická filologie - Latinská filologie	2012/2013
Anglická filologie - Německá filologie	2011/2012
Anglická filologie - Německá filologie	2012/2013
Anglická filologie - Nizozemská filologie	2011/2012
Anglická filologie - Praktická nizozemská filologie	2012/2013
Anglická filologie - Portugalská filologie	2011/2012
Anglická filologie - Portugalská filologie	2012/2013
Anglická filologie - Ruská filologie	2011/2012
Anglická filologie - Ruská filologie	2012/2013
Anglická filologie - Španělská filologie	2011/2012
Anglická filologie - Španělská filologie	2012/2013
Anglická filologie - Žurnalistika	2011/2012
Anglická filologie - Žurnalistika	2012/2013
Archeologie - Historie	2011/2012
Archeologie - Historie	2012/2013
Archivnictví - Archeologie	2012/2013
Archivnictví - Filmová věda	2012/2013
Archivnictví - Historie	2012/2013
Archivnictví - Latinská filologie	2012/2013

Archivnictví - Ruská filologie	2012/2013
Archivnictví - Žurnalistika	2012/2013
Česká filologie - Anglická filologie	2011/2012
Česká filologie - Anglická filologie	2012/2013
Česká filologie - Divadelní věda	2011/2012
Česká filologie - Filmová věda	2011/2012
Česká filologie - Latinská filologie	2011/2012
Česká filologie - Německá filologie	2011/2012
Česká filologie - Ruská filologie	2011/2012
Česká filologie - Ruská filologie	2012/2013
Česká filologie - Španělská filologie	2011/2012
Čínská filologie - Filmová věda	2012/2013
Divadelní věda - Filmová věda	2011/2012
Divadelní věda - Filmová věda	2012/2013
Divadelní věda - Historie	2011/2012
Francouzská filologie - Aplikovaná ekonomická studia	2011/2012
Francouzská filologie - Archeologie	2011/2012
Francouzská filologie - Čínská filologie	2011/2012
Francouzská filologie - Divadelní věda	2011/2012
Francouzská filologie - Filmová věda	2011/2012
Francouzská filologie - Filmová věda	2012/2013
Francouzská filologie - Historie	2011/2012
Francouzská filologie - Japonština pro hospodářskou praxi	2011/2012
Francouzská filologie - Latinská filologie	2011/2012
Francouzská filologie - Žurnalistika	2011/2012
Filozofie - Aplikovaná ekonomická studia	2011/2012
Filozofie - Aplikovaná ekonomická studia	2012/2013
Filozofie - Anglická filologie	2011/2012
Filozofie - Anglická filologie	2012/2013
Filozofie - Česká filologie	2011/2012
Filozofie - Česká filologie	2012/2013
Filozofie - Divadelní věda	2011/2012
Filozofie - Divadelní věda	2012/2013
Filozofie - Francouzská filologie	2011/2012
Filozofie - Filmová věda	2012/2013
Filozofie - Historie	2011/2012
Filozofie - Historie	2012/2013
Filozofie - Italská filologie	2011/2012
Filozofie - Italská filologie	2012/2013
Filozofie - Japonská filologie	2011/2012
Filozofie - Japonština pro hospodářskou praxi	2012/2013
Filozofie - Latinská filologie	2011/2012
Filozofie - Německá filologie	2012/2013
Filozofie - Sociologie	2012/2013

Filozofie - Španělská filologie	2012/2013
Filozofie - Žurnalistika	2011/2012
Filozofie - Žurnalistika	2012/2013
Filmová věda - Archeologie	2012/2013
Filmová věda - Historie	2011/2012
Filmová věda - Historie	2012/2013
Italská filologie - Aplikovaná ekonomická studia	2011/2012
Italská filologie - Portugalská filologie	2011/2012
Italská filologie - Ukrajinská filologie	2011/2012
Japonská filologie - Archeologie	2012/2013
Kulturní antropologie - Anglická filologie	2012/2013
Kulturní antropologie - Archeologie	2011/2012
Kulturní antropologie - Archeologie	2012/2013
Kulturní antropologie - Francouzská filologie	2011/2012
Kulturní antropologie - Filozofie	2011/2012
Kulturní antropologie - Filozofie	2012/2013
Kulturní antropologie - Filmová věda	2012/2013
Kulturní antropologie - Italská filologie	2011/2012
Kulturní antropologie - Německá filologie	2012/2013
Kulturní antropologie - Ruská filologie	2011/2012
Latinská filologie - Historie	2011/2012
Latinská filologie - Italská filologie	2011/2012
Muzikologie - Anglická filologie	2012/2013
Muzikologie - Divadelní věda	2012/2013
Muzikologie - Filozofie	2012/2013
Muzikologie - Filmová věda	2011/2012
Muzikologie - Filmová věda	2012/2013
Německá filologie - Aplikovaná ekonomická studia	2011/2012
Německá filologie - Aplikovaná ekonomická studia	2012/2013
Německá filologie - Čínská filologie	2011/2012
Německá filologie - Čínská filologie	2012/2013
Německá filologie - Filmová věda	2011/2012
Německá filologie - Historie	2011/2012
Německá filologie - Historie	2012/2013
Německá filologie - Japonská filologie	2011/2012
Německá filologie - Japonská filologie	2012/2013
Německá filologie - Japonština pro hospodářskou praxi	2011/2012
Německá filologie - Japonština pro hospodářskou praxi	2012/2013
Německá filologie - Nizozemská filologie	2011/2012
Německá filologie - Polská filologie	2011/2012
Německá filologie - Ruská filologie	2011/2012
Německá filologie - Španělská filologie	2011/2012
Německá filologie - Žurnalistika	2012/2013
Nizozemská filologie - Filmová věda	2011/2012

Praktická nizozemská filologie - Německá filologie	2012/2013
Polská filologie - Ruská filologie	2011/2012
Portugalská filologie - Divadelní věda	2011/2012
Portugalská filologie - Filmová věda	2011/2012
Ruská filologie - Aplikovaná ekonomická studia	2011/2012
Ruská filologie - Aplikovaná ekonomická studia	2012/2013
Ruská filologie - Čínská filologie	2011/2012
Ruská filologie - Čínská filologie	2012/2013
Ruská filologie - Historie	2011/2012
Ruská filologie - Japonská filologie	2012/2013
Ruská filologie - Ukrajinská filologie	2011/2012
Španělská filologie - Aplikovaná ekonomická studia	2011/2012
Španělská filologie - Čínská filologie	2011/2012
Španělská filologie - Historie	2011/2012
Španělská filologie - Japonská filologie	2011/2012
Španělská filologie - Nizozemská filologie	2011/2012
Španělská filologie - Portugalská filologie	2011/2012
Španělská filologie - Žurnalistika	2011/2012
Uměnovědná studia - Anglická filologie	2012/2013
Uměnovědná studia - Archeologie	2012/2013
Uměnovědná studia - Francouzská filologie	2011/2012
Uměnovědná studia - Německá filologie	2011/2012
Uměnovědná studia - Španělská filologie	2011/2012
Žurnalistika - Archeologie	2011/2012
Žurnalistika - Divadelní věda	2011/2012
Žurnalistika - Divadelní věda	2012/2013
Žurnalistika - Filmová věda	2011/2012
Žurnalistika - Filmová věda	2012/2013

Pozn.: S výjimkou oboru Angličtina se zaměřením na komunitní tlumočení a překlad se vždy jednalo o dvouobory a s výjimkou dvouoboru Divadelní věda - Filmová věda o prezenční studium.

Příloha č. 5: Tabulky k faktorové analýze

Tab. III: Hodnoty Kaiser-Meyer-Olkinova koeficientu pro jednotl. verze SPF 2011-2013

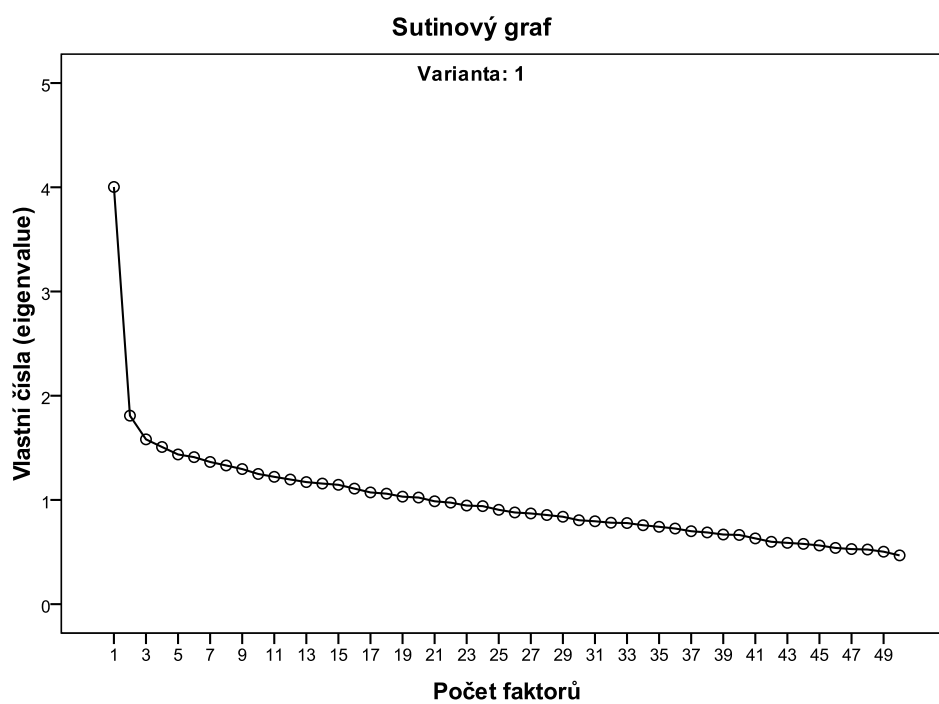
Verze testu	SPF 2011		SPF 2012		SPF 2013	
	Počet test. os.	KMO	Počet test. os.	KMO	Počet test. os.	KMO
1	584	0,70	431	0,63	367	0,58
2	558	0,75	422	0,59	374	0,58
3	333	0,67	308	0,55	375	0,55
4	336	0,66	319	0,56	362	0,58
5	257	0,57	248	0,57	125	0,48
6	245	0,62	230	0,60	130	0,48
7	100	0,41	252	0,58	344	0,61
8	93	0,37	275	0,60	624	0,69
9	357	0,68	194	0,55	342	0,57
10	335	0,71	166	0,52	644	0,68
11	360	0,60	307	0,62	276	0,58
12	340	0,60	351	0,63	268	0,55
13	345	0,65	213	0,53	X	X
14	341	0,59	196	0,50	X	X
15	151	0,47	272	0,55	X	X
16	145	0,44	275	0,57	X	X

Tab. IV: Prototyp podoby vysvětlené variance při extrakci faktorů v testech SPF 2011-2013

Faktor	Původní vlastní čísla (eigenvalues)			Vlastní čísla po extrakci		
	Celkem	% var.	kum. %	Celkem	% var.	kum. %
1	4,607	9,215	9,215	3,980	7,959	7,959
2	1,995	3,989	13,204	1,374	2,747	10,706
3	1,689	3,379	16,583	1,162	2,324	13,031
4	1,662	3,324	19,907	1,057	2,114	15,145
5	1,529	3,058	22,965	,952	1,905	17,049
6	1,448	2,896	25,861	,878	1,756	18,806
7	1,437	2,874	28,736	,814	1,628	20,434
8	1,380	2,759	31,495	,756	1,512	21,946
9	1,349	2,699	34,194	,715	1,430	23,376
10	1,334	2,667	36,861	,660	1,320	24,696
11	1,250	2,500	39,361	,648	1,295	25,991
12	1,232	2,464	41,825	,630	1,259	27,250
13	1,215	2,430	44,254	,590	1,179	28,429
14	1,193	2,386	46,640	,565	1,129	29,558
15	1,140	2,279	48,919	,478	,956	30,514
16	1,107	2,215	51,134	,468	,935	31,449
17	1,071	2,141	53,275	,448	,896	32,345
18	1,048	2,097	55,372	,408	,816	33,161
19	1,020	2,040	57,411	,382	,765	33,926
20	1,013	2,026	59,437	,368	,737	34,663
21	,974	1,949	61,386			
22	,963	1,926	63,312			
23	,927	1,854	65,166			
24	,920	1,841	67,007			
25	,894	1,788	68,795			
26	,854	1,708	70,503			
27	,847	1,693	72,196			
28	,828	1,657	73,853			
29	,814	1,629	75,482			
30	,802	1,603	77,085			
31	,773	1,546	78,630			
32	,755	1,510	80,140			
33	,741	1,482	81,621			

34	,695	1,390	83,012
35	,689	1,378	84,389
36	,677	1,354	85,743
37	,671	1,342	87,085
38	,645	1,290	88,375
39	,595	1,189	89,564
40	,576	1,152	90,716
41	,561	1,122	91,838
42	,529	1,057	92,895
43	,525	1,049	93,944
44	,507	1,014	94,959
45	,492	,984	95,943
46	,480	,959	96,902
47	,426	,851	97,753
48	,394	,787	98,540
49	,377	,755	99,295
50	,352	,705	100,000

Pozn.: Extrakce faktorů pomocí metody hlavních os; příkladem jsou vysvětlené podíly variance verze 3 z SPF 2011.



Graf I: Prototyp podoby sutinových grafů EFA testů SPF 2011-2013

Tab. V: Prototyp podoby komunalit položek a faktorů testů SPF 2011-2013

Komunality			Komunality (pokrač.)		
Položka	Původní	Po extrakci	Položka	Původní	Po extrakci
1	,165	,185	26	,222	,598
2	,200	,346	27	,164	,285
3	,190	,199	28	,141	,203
4	,202	,369	29	,252	,615
5	,215	,301	30	,196	,506
6	,180	,251	31	,225	,308
7	,222	,296	32	,184	,299
8	,177	,233	33	,260	,272
9	,160	,210	34	,208	,332
10	,210	,621	35	,254	,456
11	,239	,386	36	,200	,274
12	,191	,226	37	,269	,412
13	,267	,356	38	,345	,514
14	,271	,349	39	,137	,187
15	,189	,297	40	,208	,304
16	,181	,234	41	,246	,350
17	,220	,495	42	,257	,369
18	,200	,311	43	,124	,133
19	,322	,431	44	,150	,322
20	,207	,356	45	,169	,503
21	,091	,095	46	,268	,352
22	,225	,860	47	,183	,284
23	,144	,201	48	,313	,663
24	,242	,354	49	,196	,453
25	,109	,124	50	,142	,252

Pozn.: Extrakce faktorů pomocí metody hlavních os; příkladem jsou komunality položek verze 3 z SPF 2011 při vyextrahování 20 faktorů.

Tab. VI: Příklad podoby faktorových zátěží proměnných při rotaci 2 faktorů metodou Promax

Proměnná	Faktor 1	Faktor 2	Proměnná	Faktor 1	Faktor 2
1			31		
2			32		
3		,319	33		,332
4			34		
5			35		
6	,304		36		
7		,381	37		
8		,275	38		
9		,363	39		,321
10			40		
11			41		
12			42		
13		,284	43		,321
14			44		,322
15			45		,348
16	,306		46		
17			47		
18			48		
19	,284		49		
20			50		
21	,281				
22					
23	,363				
24	,439				
25	,330				
26	,400				
27	,328				
28	,291				
29	,287				
30					

Pozn.: Příkladem jsou faktorové zátěže (pattern matrix) položek verze 1 testu SPF 2013, zobrazovány jsou zátěže větší než 0,27. Položky 19 a 21 jsou z oddílu prostorové představivosti, položky 23-29 zahrnují téměř celý (pol.30) logicko-analytický oddíl. Naopak položky 3, 7-9 a 13 se týkají slovtvorby a frazeologie, položky 39, 43-45 otázek z literatury a historie.

Příloha č. 6: Příklad výstupu položkové analýzy

položková analýza SPF 2013 verze 01

Legend: Distractors Chosen More than Correct Answer: ■

Total Possible Points:	50,00	Median Score:	27,00	Maximum Score:	42,00
Total Students:	367	Mean Score:	27,01	Minimum Score:	9,00
Standard Deviation:	5,18	Reliability Coefficient (KR20):	0,65	Range of Scores:	33,00

No.	Question	Correct Answer	Response Frequencies					Non Distractor	Correct Group Responses			Point Biserial
			a	b	c	d	e		Total %	Upper 27%	Lower 27%	
1	pol_1	e	2,45	0,82	0,00	20,71	76,02	c	76,02	82,83	66,67	0,15
2	pol_2	b	6,27	63,49	3,81	23,98	2,18		63,49	65,66	61,62	0,04
3	pol_3	c	4,90	41,14	35,97	13,35	2,45		35,97	42,42	29,29	0,15
4	pol_4	b	9,81	9,81	20,44	23,71	35,69		9,81	12,12	11,11	0,04
5	pol_5	e	14,17	2,72	0,54	6,54	74,93		74,93	88,89	70,71	0,16
6	pol_6	d	0,54	5,72	1,91	81,47	10,08		81,47	92,93	59,60	0,32
7	pol_7	b	3,00	79,84	7,36	1,36	8,45		79,84	97,98	57,58	0,38
8	pol_8	d	0,27	1,09	1,36	96,73	0,54		96,73	100,00	88,89	0,29
9	pol_9	b	12,53	54,22	18,53	4,63	7,36		54,22	78,79	29,29	0,41
10	pol_10	b	0,27	77,93	4,36	15,26	1,91		77,93	88,89	72,73	0,18
11	pol_11	b	8,17	68,39	10,90	7,36	4,36		68,39	80,81	68,69	0,12
12	pol_12	a	66,49	3,27	1,36	3,54	24,25		66,49	79,80	45,45	0,28
13	pol_13	a	59,67	7,63	32,15	0,27	0,00	e	59,67	81,82	34,34	0,34
14	pol_14	b	20,98	31,88	2,72	30,25	9,54		31,88	38,38	34,34	0,07
15	pol_15	c	5,18	6,81	62,40	10,08	14,17		62,40	81,82	44,44	0,29
16	pol_16	b	1,36	42,51	1,91	8,17	45,50		42,51	63,64	23,23	0,34
17	pol_17	d	0,00	6,81	6,81	81,20	4,63	a	81,20	90,91	69,70	0,23
18	pol_18	a	79,29	3,00	7,08	5,45	3,00		79,29	88,89	65,66	0,25
19	pol_19	c	3,54	9,54	59,95	20,16	4,90		59,95	71,72	51,52	0,15
20	pol_20	c	1,09	4,09	84,74	3,00	4,63		84,74	87,88	72,73	0,18
21	pol_21	b	12,53	55,59	9,26	13,08	8,99		55,59	70,71	37,37	0,23
22	pol_22	c	11,44	20,71	17,71	20,16	19,35		17,71	28,28	14,14	0,17
23	pol_23	c	5,99	7,63	74,66	6,27	2,18		74,66	85,86	63,64	0,25
24	pol_24	e	14,17	2,45	21,80	14,71	45,23		45,23	64,65	21,21	0,35
25	pol_25	a	52,04	7,36	5,72	16,62	10,90		52,04	67,68	35,35	0,23

Obr.II.: Příklad výstupu položkové analýzy testů SPF 2011-2013.

Příloha č. 7: Výsledky regresní analýzy pro H5a-b

Tab. VII: Hodnoty regresních koeficientů a statistická významnost příslušného modelu v krokové regresní analýze pro ověření H5a a H5b

Ak. rok	Krok	Prediktory	Regresní koeficienty				ANOVA pro model							
			B	SDCh	Beta	t	p	F	sv1	sv2	p			
2011/2012	1	konstanta	1,580	0,051		31,147	0,000	53,889	1	795	0,000			
		mat.průměr	,203	0,028	,252	7,341	0,000							
	2	konstanta	1,918	0,139		13,784	0,000		30,537			2	794	0,000
		mat.průměr	,179	0,029	,223	6,196	0,000							
		body SPF	-,005	0,002	-,094	-2,606	0,009							
	3	konstanta	2,358	0,142		16,613	0,000		48,246			3	793	0,000
		mat.průměr	,234	0,028	,290	8,245	0,000							
		body SPF	-,005	0,002	-,098	-2,859	0,004							
		věk	-,023	0,003	-,296	-8,818	0,000							
2012/2013	1	konstanta	1,605	0,060		26,958	0,000	90,661	1	957	0,000			
		mat.průměr	,312	0,033	,294	9,522	0,000							
	2	konstanta	2,588	0,147		17,601	0,000		74,259			2	956	0,000
		mat.průměr	,244	0,033	,230	7,334	0,000							
		body SPF	-,015	0,002	-,228	-7,276	0,000							
	3	konstanta	2,639	0,160		16,484	0,000		49,710			3	955	0,000
		mat.průměr	,249	0,034	,235	7,358	0,000							
		body SPF	-,014	0,002	-,226	-7,199	0,000							
		věk	-,003	0,004	-,025	-,815	0,415							

Pozn.: B = hodnota nestandardizovaného regresního koeficientu; SDCh = standardní chyba odhadu; Beta = hodnota standardizovaného regresního koeficientu; sv1 a sv2 = stupně volnosti.

Příloha č. 8: Výsledky regresní analýzy pro H6a a H6c

Tab. VIII: Hodnoty regresních koeficientů a statistická významnost příslušného modelu v krokové regresní analýze pro ověření H6a a H6c

Ak. rok	Krok	Prediktory	Regresní koeficienty					ANOVA pro model				
			B	SDCh	Beta	t	p	F	sv1	sv2	p	
2011/2012	1	konstanta	1,873	0,125		14,972	0,000	0,889	1	343	0,346	
		OT/ÚZ aj.	,002	0,002	,051	,943	0,346					
	2	konstanta	1,911	0,200		9,548	0,000	0,473	2	342	0,623	
		OT/ÚZ aj.	,002	0,002	,051	,941	0,347					
		body SPF	-,002	0,009	-,013	-,244	0,808					
	3	konstanta	1,558	0,243		6,419	0,000	2,459	3	341	0,063	
		OT/ÚZ aj.	,003	0,002	,071	1,311	0,191					
		body SPF	,004	0,009	,026	,474	0,636					
		mat.průměr	,108	0,043	,143	2,533	0,012					
2012/2013	1	konstanta	3,254	0,135		24,093	0,000	77,135	1	431	0,000	
		OT/ÚZ aj.	-,022	0,003	-,390	-8,783	0,000					
	2	konstanta	4,277	0,164		26,150	0,000	91,531	2	430	0,000	
		OT/ÚZ aj.	-,017	0,002	-,295	-7,095	0,000					
		body SPF	-,074	0,008	-,395	-9,487	0,000					
	3	konstanta	3,803	0,210		18,131	0,000	66,830	3	429	0,000	
		OT/ÚZ aj.	-,015	0,002	-,265	-6,336	0,000					
		body SPF	-,068	0,008	-,362	-8,599	0,000					
		mat.průměr	,157	0,044	,149	3,539	0,000					

Pozn.: B = hodnota nestandardizovaného regresního koeficientu; SDCh = standardní chyba odhadu; Beta = hodnota standardizovaného regresního koeficientu; sv1 a sv2 = stupně volnosti.

Příloha č. 9: Výsledky regresní analýzy pro H6b a H7b

Tab. IX: Hodnoty regresních koeficientů a statistická významnost příslušného modelu v krokové regresní analýze pro ověření H6b

Krok	Prediktory	Regresní koeficienty					ANOVA pro model				
		B	SDCh	Beta	t	p	F	sv1	sv2	p	
1	konstanta	1,375	1,079		1,275	0,206	0,001	1	88	0,971	
	OT/ÚZ aj.	,001	0,018	,004	0,036	0,971					
2	konstanta	1,591	1,156		1,376	0,172	0,144	2	87		
	OT/ÚZ aj.	,001	0,018	,007	0,066	0,948				0,866	
	body SPF	-,010	0,019	-,057	-0,535	0,594					
3	konstanta	1,087	1,103		0,985	0,327	3,880	3	86	0,012	
	OT/ÚZ aj.	,002	0,017	,014	0,135	0,893					
	body SPF	-,007	0,018	-,038	-0,370	0,712					
	mat.průměr	,222	0,066	,341	3,364	0,001					

Tab. X: Hodnoty regresních koeficientů a statistická významnost příslušného modelu v krokové regresní analýze pro ověření H7b

Krok	Prediktory	Regresní koeficienty					ANOVA pro model				
		B	SDCh	Beta	t	p	F	sv1	sv2	p	
1	konstanta	1,061	0,109		9,780	0,000	11,736	1	88	0,001	
	mat.průměr	,224	0,065	,343	3,426	0,001					
2	konstanta	,943	1,027		0,918	0,361	5,809	2	87	0,004	
	mat.průměr	,224	0,066	,343	3,408	0,001					
	OT/ÚZ aj.	,002	0,017	,012	0,116	0,908					
3	konstanta	1,087	1,103		0,985	0,327	3,880	3	86	0,012	
	mat.průměr	,222	0,066	,341	3,364	0,001					
	OT/ÚZ aj.	,002	0,017	,014	0,135	0,893					
	body SPF	-,007	0,018	-,038	-,370	0,712					
4	konstanta	1,272	1,092		1,165	0,247	3,892	4	85	0,006	
	mat.průměr	,213	0,065	,327	3,264	0,002					
	OT/ÚZ aj.	-,003	0,017	-,017	-0,169	0,866					
	body SPF	-,011	0,018	-,060	-0,600	0,550					
	věk	,010	0,005	,193	1,892	0,062					

Příloha č. 10: Výsledky regresní analýzy pro H7a a H7c

Tab. XI: Hodnoty regresních koeficientů a statistická významnost příslušného modelu v krokové regresní analýze pro ověření H7a a H7c

Ak. rok	Krok	Prediktory	Regresní koeficienty					ANOVA pro model			
			B	SDCh	Beta	t	p	F	sv1	sv2	p
2011	1	konstanta	1,818	0,077		23,490	0,000	5,485	1	343	0,020
		mat.průměr	,095	0,041	,125	2,342	0,020				
	2	konstanta	1,647	0,154		10,723	0,000	3,584	2	342	0,029
		mat.průměr	,102	0,041	,135	2,503	0,013				
		OT/ÚZ aj.	,003	0,002	,070	1,293	0,197				
	3	konstanta	1,558	0,243		6,419	0,000	2,459	3	341	0,063
		mat.průměr	,108	0,043	,143	2,533	0,012				
		OT/ÚZ aj.	,003	0,002	,071	1,311	0,191				
		body SPF	,004	0,009	,026	0,474	0,636				
	4	konstanta	2,306	0,242		9,541	0,000	18,230	4	340	0,000
		mat.průměr	,193	0,041	,255	4,749	0,000				
		OT/ÚZ aj.	-,003	0,002	-,074	-1,392	0,165				
body SPF		,008	0,008	,052	1,018	0,309					
věk		-,027	0,003	-,439	-8,011	0,000					
2012	1	konstanta	1,514	0,090		16,832	0,000	46,694	1	431	0,000
		mat.průměr	,328	0,048	,313	6,833	0,000				
	2	konstanta	2,657	0,175		15,170	0,000	54,100	2	430	0,000
		mat.průměr	,241	0,047	,229	5,148	0,000				
		OT/ÚZ aj.	-,019	0,003	-,332	-7,456	0,000				
	3	konstanta	3,803	0,210		18,131	0,000	66,830	3	429	0,000
		mat.průměr	,157	0,044	,149	3,539	0,000				
		OT/ÚZ aj.	-,015	0,002	-,265	-6,336	0,000				
		body SPF	-,068	0,008	-,362	-8,599	0,000				
	4	konstanta	4,072	0,222		18,312	0,000	54,112	4	428	0,000
		mat.průměr	,178	0,044	,170	4,030	0,000				
		OT/ÚZ aj.	-,017	0,002	-,291	-6,906	0,000				
		body SPF	-,064	0,008	-,339	-8,058	0,000				
		věk	-,013	0,004	-,137	-3,346	0,001				