

INTERPRETACE VÝZKUMNÝCH OTÁZEK RELIABILITY JAZYKOVÝCH TESTŮ

Hana Romová¹; Tomáš Římský²

Policejní akademie ČR v Praze, Fakulta bezpečnostního managementu, Katedra jazyků,
Lhotecká 559/7, 143 00 Praha 4, Česká republika
e-mail: romova@polac.cz; tomas.rimsky@polac.cz

Abstrakt

Tento článek řeší problematiku reliability jazykových didaktických testů v rámci jazykového testování. Získaná data slouží pro interpretaci a moderaci testových položek na katedře jazyků. Testování byli studenti 1 a 2. ročníků Fakulty bezpečnostního managementu a Fakulty bezpečnostně právní Policejní akademie ČR v Praze, oborů Bezpečnostní management ve veřejné správě a Bezpečnostně právní studia. Pro stanovení reliability byl stanoven jako měřitelný indikátor index vnitřní konzistence s rozdělením do škál, které byly použity pro stanovení kritéria pro hodnocení reliability sub-testů. Reliabilita sub-testů je interpretována v širším kontextu a je důležitým krokem k optimalizaci testovacích procesů v rámci vytváření a doplňování testové banky pro katedru jazyků.

Keywords

Test construction; Items; Reliability; Distractors; Terms; Terminology; Practice.

Úvod

V rámci testování a vyhodnocování prací studentů na katedře jazyků jsme přistoupili k implementaci digitalizovaných dat získaných z výstupů jazykových testů.

Data získaná ve výzkumu, která byla generována v procesu testování, poskytují testerům objektivnější základ pro interpretaci a aplikaci získaných výstupů a jsou v tomto smyslu nenahraditelným vodítkem pro interpretaci a následnou moderaci testových položek a při hledání odpovědí na výzkumné otázky. Jedním z těchto kritérií, na které jsme se zaměřili v této části výzkumu, jsou otázky týkající se reliability jazykových testů a vzájemné srovnatelnosti reliability testů prověřujících jednotlivé jazykové dovednosti [1]. Pod termínem reliability testu si můžeme představit měřenou či odhadovanou hodnotu, která představuje míru, do jaké dokážeme testovat konzistentně a přesně.

1 Metody výzkumu

Při testování je třeba vycházet ze tří základních otázek: *co bude hodnoceno, jak*, tj. jaké nástroje a způsob „měření“ je třeba zvolit, a *v jakém kontextu*, tj. mimo jiné k jakému účelu budou výsledky testování sloužit.

Jazykové testy mají nepochybně dlouhou historii. Za první odbornou monografii k problematice testování (anglického) jazyka bývá považováno *Language Testing*, které R. Lado publikoval v roce 1961. Koncem 70. a počátkem 80. let 20. století se začíná rozvíjet jeden z dosud nevlivnějších trendů, komunikační přístup k jazykovému testování [5].

V návaznosti na rozbor skutečných potřeb testování byly především pro položkovou banku pro anglický jazyk vypracovány zásady jejího vytváření, konstruktů a specifikací testů a jejich obsahové náplně vycházející z učebnic doporučených pro studium [2]. Položková banka, terminologicky též nazývaná banka testových úloh (BTÚ) nebo banka testových položek

(BTP), anglicky Item Bank, v našem pojetí operuje s jednotlivými kompletními svazky položek, které jsou výsledkem promyšleného sestavování tvůrcem testu, nikoli náhodného generování počítačem či náhodného výběru tvůrcem testu ze sumy položek. Vzhledem k našim technickým možnostem vytváříme tuto banku ve smyslu úložiště testových položek, nikoli ve smyslu celého informačního systému se všemi procesy až po vygenerování testu počítačem. Jako metody výzkumu byly použity **receptivní testovací techniky**. K výzkumným otázkám bylo provedeno empirické šetření ilustrativního charakteru, dílčí statistické vyhodnocení a interpretace některých získaných dat.

2 Předmět výzkumu

Předmětem našeho výzkumného zájmu byla vybraná problematika reliability jazykových didaktických testů [1].

Z několika běžně používaných přístupů ke generaci dat pro stanovení reliability testů jsme zvolili reliability akcentující vnitřní konsistenci testu. Tento postup je založen na jedné administraci jednoho testu jedné skupině respondentů. Indikátor reliability pak ukazuje míru, do jaké sledované položky testu měří konzistentně stejnou věc (např. porozumění psanému textu, kdy respondent na položky z jedné oblasti (sub-testu) odpovídá podobně a odpovědi pak spolu korelují) [7].

Jako číselný měřitelný indikátor jsme použili index Cronbachova alfa, který ukazuje na korelaci skóre všech položek testu. Tento index, který bývá někdy označován jako index vnitřní konzistence, se pohybuje v hodnotách od 0 do 1 s rozdělením do škál odpovídajícím různé míře konsistence testu. Tyto škály následovně používáme jako kritérium pro hodnocení reliability sub-testů, viz Tab. 1.

Tab. 1: Indikátor alfa a interpretační škála použita při výzkumu

Cronbachova alfa	Míra vnitřní konsistence testu
$\alpha \geq 0.9$	Výborná
$0.9 > \alpha \geq 0.8$	Dobrá
$0.8 > \alpha \geq 0.7$	Přípustná
$0.7 > \alpha \geq 0.6$	Problématická
$0.6 > \alpha \geq 0.5$	Špatná
$0.5 < \alpha$	Nevyhovující

Zdroj: [4]

Při interpretaci jsme si však také vědomi skutečnosti, že při hodnocení ukazatele Cronbachova alfa je nutné přihlédnout k okolnostem, za kterých bylo testování realizováno. Platí zde obecná závislost, že pro vyšší relevanci indikátoru Cronbachova alfa je vhodnější menší počet položek a větší počet respondentů. S nárůstem počtu testovaných položek úměrně narůstá i hodnota Cronbachova alfa, čímž tento indikátor ztrácí na relevanci pro další interpretaci. Stejně pak platí, že Cronbachova alfa je relevantním ukazatelem pro jedno-faktorové testy, přičemž pro více-faktorové testy je lépe využitelný indikátor McDonaldovo omega. Při analýze se proto neomezujeme pouze na jednu statistickou hodnotu, ale hodnotíme vždy sub-testy v kontextu s přihlédnutím k dalším faktorům, např. k vzorku respondentů nebo faktorovosti testu [6].

Tab. 2 uvádí výstupní data z realizovaných testů relevantní pro posuzování testové reliability.

Tab. 2: Výstup z datové analýzy

Skupina / Semestr TEST	Alfa	Omega	Počet respondentů
B60 1. semestr			
Listening	0,3657	0.258679	200
Reading	0,8087	0.804603	
Vocabulary	0,963	0.516512	
Grammar	0,9395	0.781337	
B60 3. semestr			
Listening	0,8892	0.434316	134
Reading	0,9345	0.929683	
Vocabulary	0,9632	0.966356	
Grammar	0,9542	0.924319	
B71 1. semestr			
Listening	0,7325	0.619843	87
Reading	0,1988	0.521898	
Vocabulary	0,8046	0.004019	
Grammar	0,8475	0.591863	
B71 3. semestr			
Listening	0,6684	0.539776	76
Reading	0,8173	0.632205	
Vocabulary	0,9071	0.157758	
Grammar	0,9171	0.302243	
B71 4. semestr			
Listening	0,6177	1.045658	72
Reading	0,7374	0.747673	
Vocabulary	0,9329	0.268062	
Grammar	0,8668	0.644990	

Zdroj: Vlastní

B60 označuje Bezpečnostně právní studia a B71 Bezpečnostní management ve veřejné správě. Pro přehlednější hodnocení výsledků testů podle indikátoru reliability je z praktického hlediska vhodné testy rozdělit do dvou skupin podle velikosti alfa, a následně se zaměřit na celkovou interpretaci dat. Interpretaci provádíme vždy pro jednotlivou skupinu sub-testu podle zaměření, viz Tab. 3.

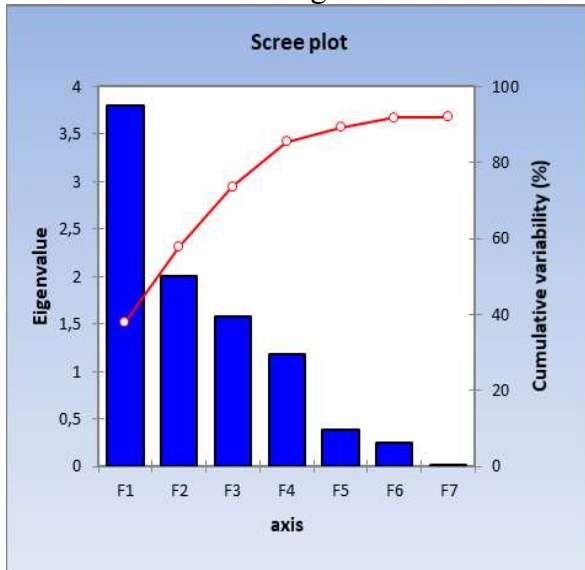
3 Reliabilita sub-testu Listening

Tab. 3: Hodnoty alfa – subtest Listening

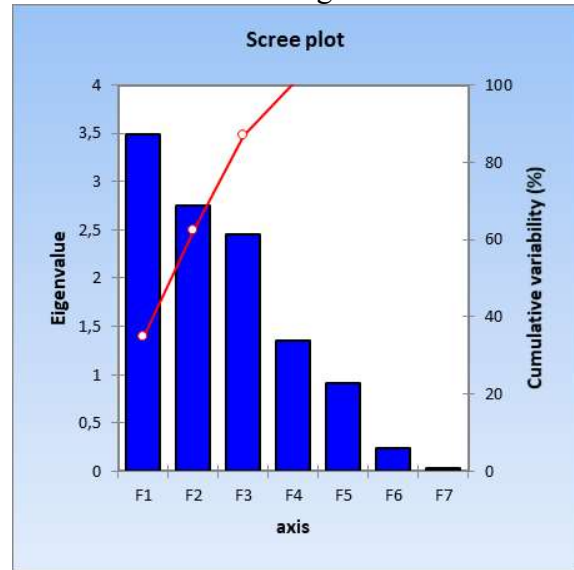
Sub-test	B60 1. semestr Alfa	B60 3. semestr Alfa	B71 1. semestr Alfa	B71 3. semestr Alfa	B71 4. semestr Alfa	Průměr Alfa
Listening	0,3657	0,8892	0,7325	0,6684	0,6177	0,6547

Zdroj: Vlastní

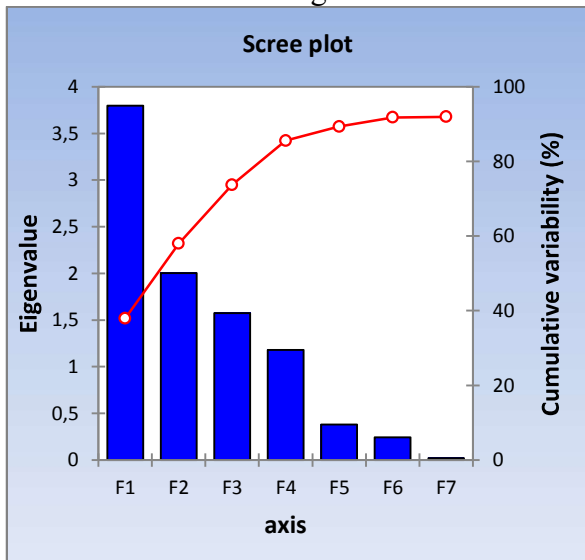
B60 1. semestr Listening



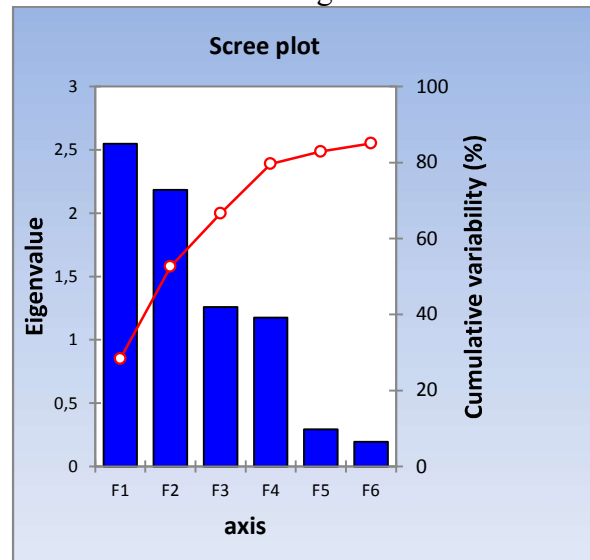
B60 3. semestr Listening



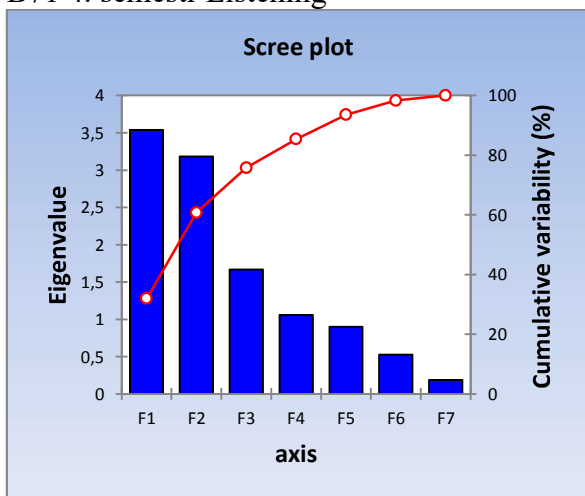
B71 1. semestr Listening



B71 3. semestr Listening



B71 4. semestr Listening



Zdroj: Vlastní

Fig. 1: Faktorová zátěž – sub-testy Listening

Výsledky sub-testu Listening s průměrnou úrovní alfa 0,6547 vykazují hodnoty na hodnotící škále – dobrá, akceptovatelná, slabá – s výjimkou testu B60 1. semestr, jehož hodnota je alfa neakceptovatelná a tento test se zdá být nejproblematictější. Kromě tohoto problematického testu je zde tedy možno vysledovat jistou konzistenci výsledků s tím, že hodnoty reliability jsou v průměru horší než u zbývajících tří sub-testů. Tento fakt lze vysvětlit tendencí k více-faktorovosti u sub-testu Listening, viz obrázek 1 (Fig. 1).

Vzhledem k výše zmíněné tendenci k více-faktorovosti u sub-testů Listening lze konstatovat, že indikátor reliability alfa má pouze informativní charakter. U více-faktorových testů má větší vypovídací hodnotu indikátoru omega, viz Tab. 2. Nicméně i za použití indexu omega zůstávají výsledky Listening sub-optimální.

Dalším aspektem vysvětlujícím již zmiňovanou tendenci relativní obtížnost dovednosti Listening, která může způsobovat asymetrie v odpovědích různých studentů, které pak následně vedou k nižšímu koeficientu alfa. Tato zvýšená obtížnost sub-testu Listening je obzvláště patrná mezi studenty kombinovaného studia, na jejichž menším úspěšném skóre se zcela logicky projevuje výrazně nižší dotace výukových hodin včetně konzultací v porovnání se studenty prezenčního studia. Lze se tedy domnívat, že studenti kombinovaného studia častěji využívají možnosti odpovídat náhodně, což ve svém důsledku znamená, že se snižuje relevance koeficientu alfa.

Na základě uvedeného lze konstatovat, že pro hodnocení reliability, testy Listening dosahují ve většině případů srovnatelných skóre z hlediska jednotlivých studijních oborů. V následujícím období je třeba dále vyladit stávající sub-testy, obzvláště ty které mají nižší skóre reliability. Je také na zvážení a na případném experimentálním prověření, zdali zvolená struktura sub-testu, (a) 5 bodů true/fause, (b) 5 bodů multiple-choice na bázi dvou různých nahrávek, nevede k více-faktorovosti a tím i k horším výsledkům indexu reliability alfa.

Při tvorbě sub-testů Listening je třeba dále sledovat indikátory reliability a věnovat více pozornosti distraktorům, které ovlivňují vnitřní konzistenci testu. Ověřili jsme si, že nastavením procesů moderace lze docílit zkvalitnění sub-testů Listening, a tím i zvýšení objektivit testů. Z pohledu optimalizace testování je relevantní soustředění se na problematictější sub-testy a aplikaci zásad pro kvalitní přípravu testovacího cyklu (sestavení testu, pilotní testace, moderace, re-testace, další modifikace testu) a s tímto výhledem je integrovat do dalšího testovacího období [3].

4 Reliabilita sub-testu Reading

Tab. 4: Hodnoty Alfa – subtest Reading

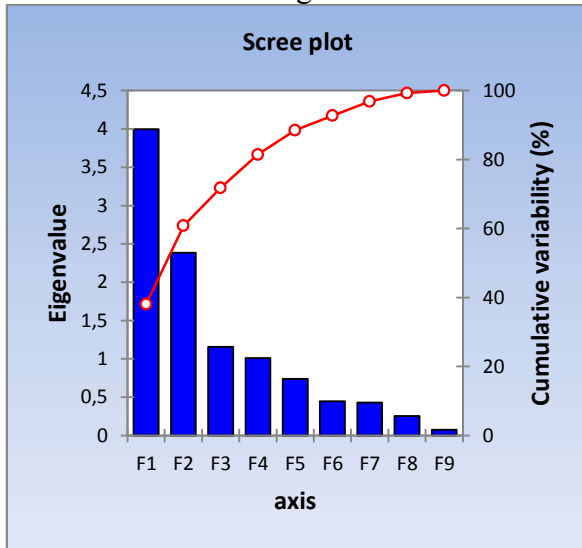
Sub-test	B60 1. semestr Alfa	B60 3. semestr Alfa	B71 1. semestr Alfa	B71 3. semestr Alfa	B71 4. semestr Alfa	Průměr Alfa
Reading	0,8087	0,9345	0,1988	0,8173	0,7374	0,6933

Zdroj: Vlastní

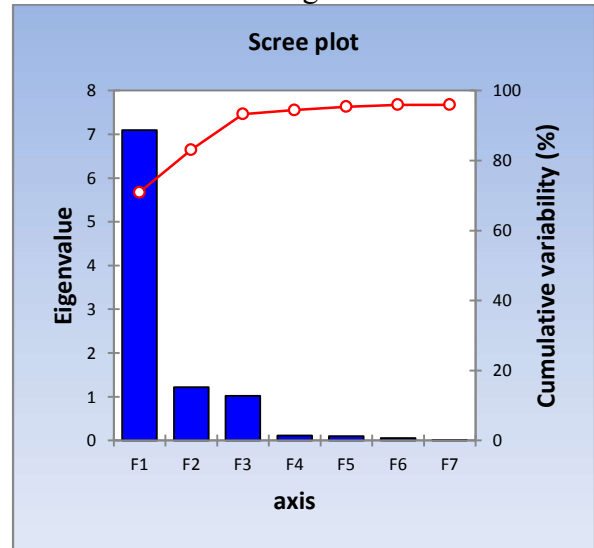
Sub-testy Reading vykazují průměrně lepší hodnoty reliability ve srovnání s předchozí skupinou taktéž deseti-položkových sub-testů Listening. Vyjma testu B71 1.sem., který sráží celkové skóre, přesáhlo alfa ve šech testech hodnotu 0,7, viz Tab. 4. Lze tedy konstatovat, že testy mají *přípustnou* úroveň reliability.

Na grafu faktorové analýzy problematického testu B71 1. semestr, viz obrázek 2 (Fig. 2), je na první pohled zcela patrné, že v porovnání s ostatními testy je silně vícedimenzionální. Reliabilitu u testů s tendencí k více-dimenzionalitě je pak lépe indikovat z indexu omega, které v tomto případě činí 0.521898, což je podstatně vyšší hodnota než index alfa.

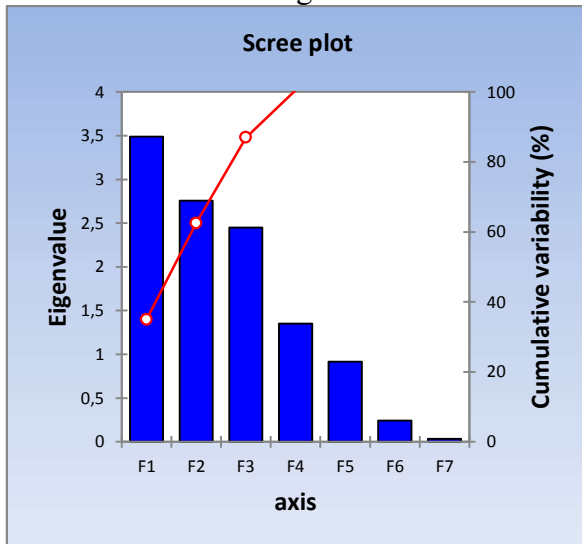
B60 1. semestr Reading



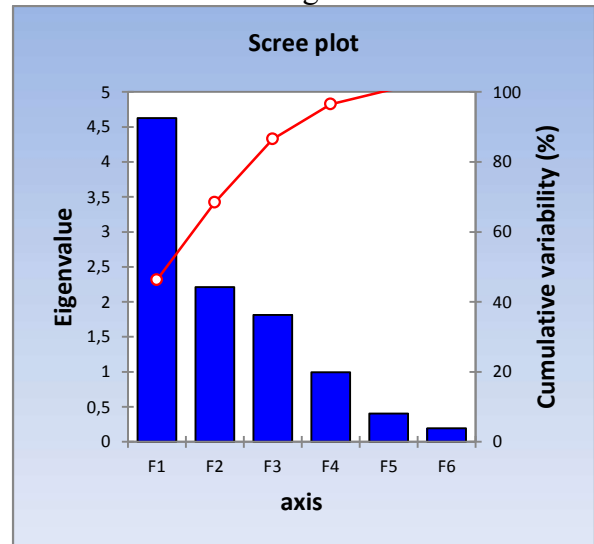
B60 3. semestr Reading



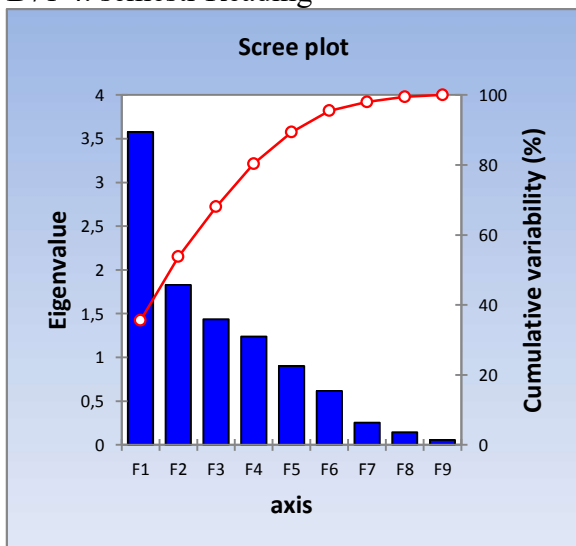
B71 1. semestr Reading



B71 3. semestr Reading



B71 4. semestr Reading



Zdroj: Vlastní

Fig. 2: Faktorová zátěž – sub-testy Reading

Nicméně více-dimenzionalita je zde z principu problematická i z jiného důvodu. V sub-testech Listening a Reading testujeme vždy jen jednu dovednost. Bylo by tedy žádoucí, aby sub-test testoval pouze jednu oblast, čímž by byl vnitřně konzistentní. V případě variant sub-testů s nízkým skóre indexu alfa je nutné provést následnou položkovou analýzu a pojmenovat a nahradit položky, které sytí jiný faktor, než je faktor dominantní. Pro identifikaci položek se osvědčilo použití numerických nebo grafických výstupů z analýzy dat.

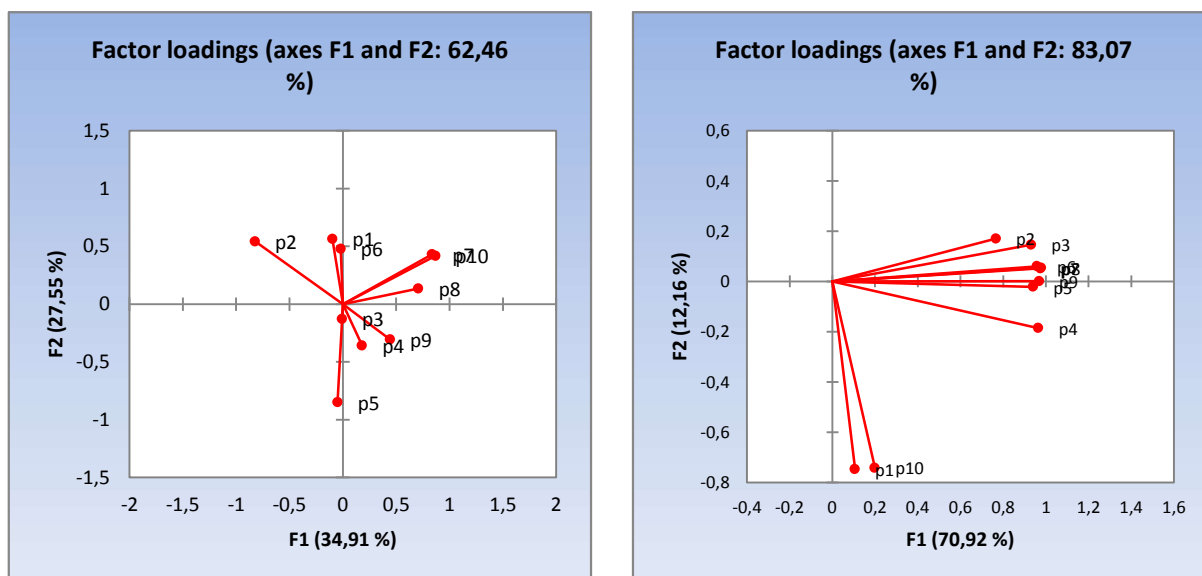
Tab. 5: Identifikace problémových položek pro test B71 3. semestr

	F1	F2	F3	F4	F5	Initial communality	Final communality	Specific variance
p1	-0,0973	0,5640	0,0915	-0,8108	0,0816	0,4771	1,0000	0,0000
p2	-0,8221	0,5416	-0,0282	-0,1637	0,0566	1,2822	1,0000	0,0000
p3	-0,0065	-0,1289	-0,1275	-0,4064	0,2140	0,3377	0,2438	0,7562
p4	0,1811	-0,3584	0,8680	-0,2710	-0,1088	1,7927	1,0000	0,0000
p5	-0,0496	-0,8501	-0,2044	-0,3935	-0,2797	1,3389	1,0000	0,0000
p6	-0,0191	0,4772	-0,5000	-0,0687	-0,5035	0,5559	0,7364	0,2636
p7	0,8373	0,4299	0,3173	-0,0737	-0,0894	0,9551	1,0000	0,0000
p8	0,7086	0,1340	-0,5937	-0,1461	-0,3258	0,8406	1,0000	0,0000
p9	0,4432	-0,3068	-0,6287	-0,1935	0,5261	0,6953	1,0000	0,0000
p10	0,8683	0,4169	0,2174	0,0413	0,1524	0,9234	1,0000	0,0000

Tučně vytištěné hodnoty indikují, že položka „sytí“ faktor v záhlaví sloupce.

Zdroj: Vlastní

Bereme-li v úvahu jako nosný faktor 1 (F1 v Tab. 5), pak by položky, které „sytí“ jiné faktory měly být modifikovány. Jedná se o položky 1, 3, 4, 5, 6, 9. Z toho vyplývá, že v případě problematických sub-testů, jako je tento případ, je nutná celková rekonstrukce položkového svazku, protože vnitřní konzistence sub-testu je ze statistického pohledu neuspokojivá. Graficky je tento stav názorně vyjádřen grafem na obrázku 3 (Fig. 3), kde u problémového sub-testu (Reading B71 3. semestr) jsou položky rozptýleny namísto toho, aby spolu do značné míry korelovaly, viz sub-test B60 3. semestr.



Zdroj: Vlastní

Fig. 3: Porovnání sub-testu B71 3. semestr s nízkou reliabilitou a sub-testu B60 3. semestr s vysokou reliabilitou

Z výše uvedeného vyplývá závěr pro stanovení reliability sub-testů Reading. Reliabilita čtyř sub-testů je dobrá, ovšem jeden sub-test neprošel kritériem reliability, a proto musí být přepracován nebo nahrazen. V rámci řešení si vědecký tým na vzorku pro pilotní testování vyzkoušel a nastavil odborné postupy, které umožňují na základě tvrdých dat i expertního posouzení testerů eliminovat takové sub-testy Reading, které jsou vnitřně nekonzistentní a jejichž výsledky by mohly být problematické.

5 Reliabilita sub-testu Vocabulary

Tab. 6: Hodnoty Alfa - Subtest Vocabulary

Sub-test	B60 1. semestr Alfa	B60 3. semestr Alfa	B71 1. semestr Alfa	B71 3. semestr Alfa	B71 4. semestr Alfa	Průměr Alfa
Vocabulary	0,9630	0,9632	0,8046	0,9071	0,9329	0,9080

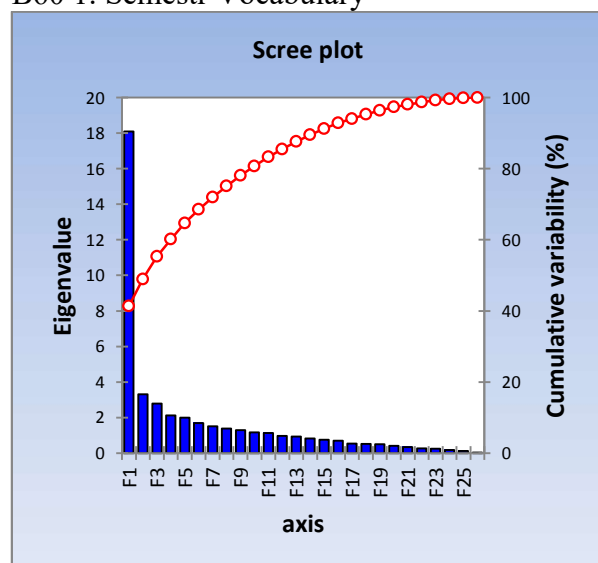
Zdroj: Vlastní

Ukazatele reliability alfa u sub-testu jazykových prostředků Vocabulary, viz Tab. 6, je podle škály *dobrá* (sub-test B71 1. semestr) a v případě čtyř dalších sub-testů *výborná*. Tento fakt je pozitivní a ukazuje na vysokou vnitřní konzistenci sub-testů v této kategorii.

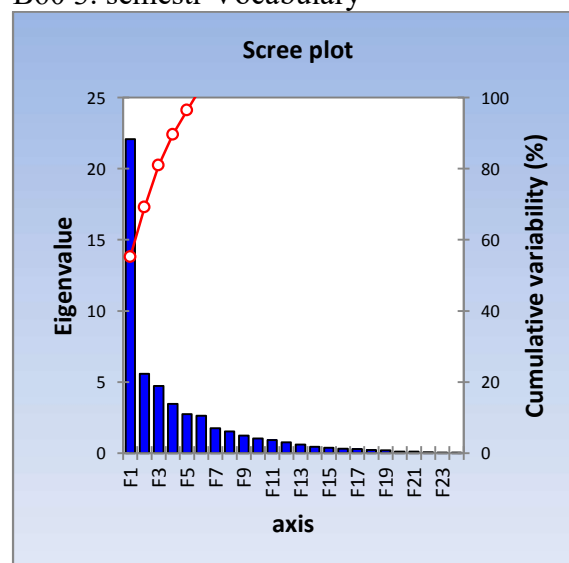
Nicméně při kritickém pohledu na dané výsledky v kontextu konstrukce i administrace sub-testů, existují skutečnosti, které mohou umenšovat váhu výše uvedených výstupů. Jako první si bedlivý pozorovatel povšimne skutečnosti, že se zde jedná o 40ti-položkový sub-test. Sama tato skutečnost statisticky zvyšuje výsledné hodnoty alfa a tím ubírá na relevanci indexu. Stejně tak působí skutečnost, že položkově relativně bohatý sub-test pracuje s malým počtem respondentů. Pro tento fakt sice existují objektivní důvody, a to maximální kapacity jednotlivých skupin, nicméně stálo by za úvahu v příštích testováních a měřeních upravit poměr mezi počtem položek a počtem respondentů, a to tak, aby poměr byl alespoň 1:10. Prakticky by bylo možné toho dosáhnout rozdělením 40ti-položkové sady na menší svazky, např. po deseti položkách, a ty by se následně vyhodnocovaly separátně. Měření po rozdělení by mohla být přesnější a více reflektovat kvalitu každé jednotlivé položky. Výsledky by potom také usnadnily položkovou analýzu sub-testu a jejich další moderaci.

Z hlediska četnosti faktorů (viz. níže) je zřejmé, že testy mají tendenci k jedno-faktorovosti. Dominance jednoho klíčového faktoru je ve většině případů zřetelná. Tento závěr posiluje předchozí zjištění o reliabilitě sub-testů měřené indexem alfa.

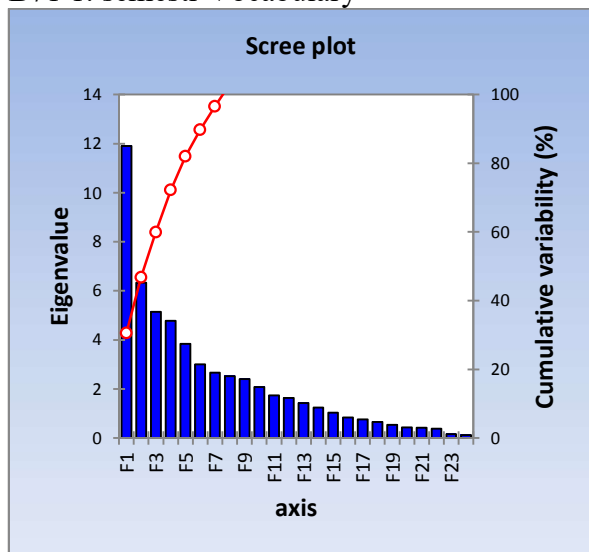
B60 1. Semestr Vocabulary



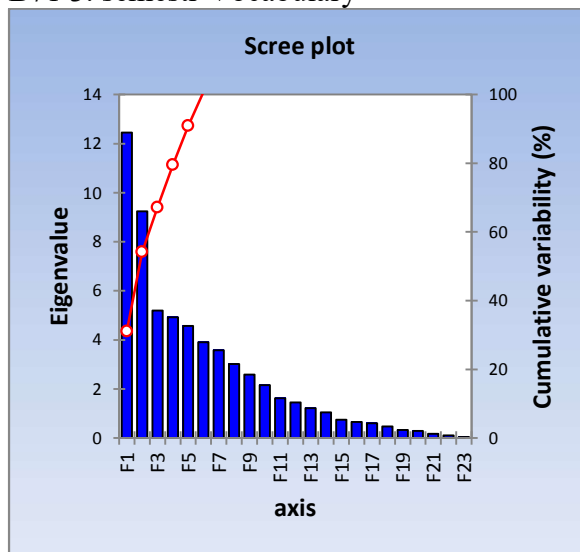
B60 3. semestr Vocabulary



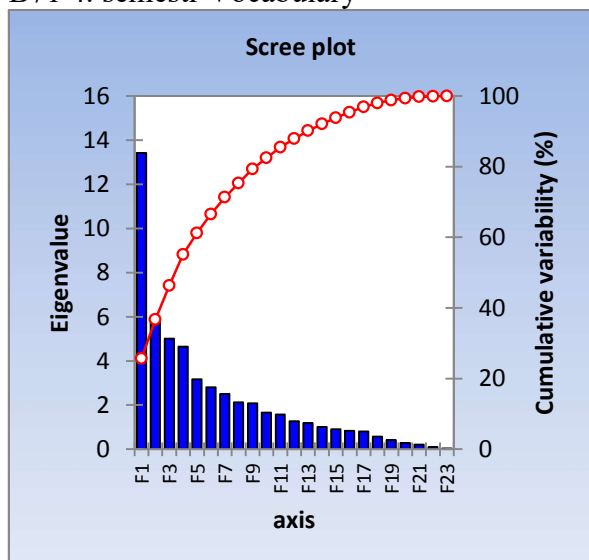
B71 1. semestr Vocabulary



B71 3. semestr Vocabulary



B71 4. semestr Vocabulary



Zdroj: Vlastní

Fig. 4: Faktorová zátěž - sub-testy Vocabulary

Závěr z analýzy reliability sub-testu Vocabulary sumarizuje, že sledovaný indikátor reliability dosahuje s přihlédnutím k výše zmíněných poznámka požadovaných hodnot, a to jak, pokud jde o jeho číselné vyjádření (alfa), tak průřezově na vzorku pěti pilotních sub-testů, kde naměřené hodnoty jsou konzistentní a vzájemně vyvážené, viz obrázek 4 (Fig. 4). Vzhledem k tomu, že práce s odbornou slovní zásobou představuje klíčovou část výuky cizího jazyka, jsou zjištění, že generované položkové testovací soubory (sub-testy) představují relevantní a spolehlivý nástroj pro hodnocení výstupních znalostí respondentů, nepochybně pozitivním signálem.

6 Reliabilita sub-testu Grammar

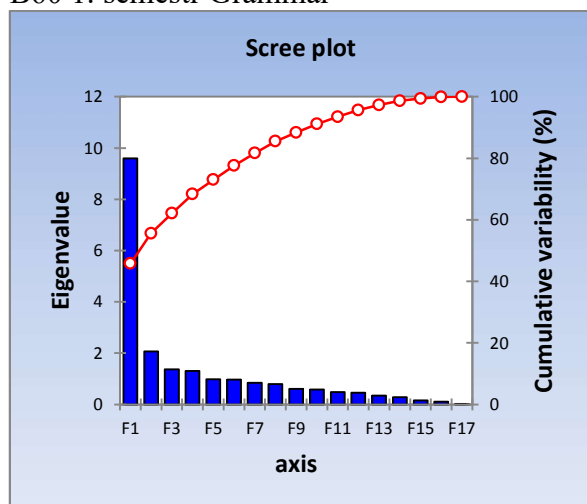
Tab. 7: Hodnoty Alfa – Subtest Grammar

Sub-test	B60 1. semestr Alfa	B60 3. semestr Alfa	B71 1. semestr Alfa	B71 3. semestr Alfa	B71 4. semestr Alfa	Průměr Alfa
Grammar	0,9395	0,9542	0,8475	0,9171	0,8668	0,9050

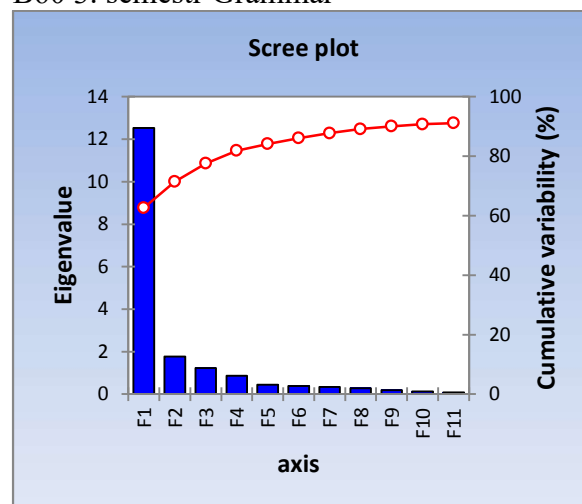
Zdroj: Vlastní

Hodnoty reliability testované na souboru pěti pilotních sub-testů Grammar, viz Tab. 7, dosahují konstantně vysokých hodnot s výslednou průměrnou hodnotou 0,9050, která odpovídá na škále hodnocení alfa klasifikaci – *výborná*. Tato skutečnost je pozitivní i v kontextu množství testovaných položek (sub-test Grammar má 20 položek), a to i přesto, že poměr počtu položek v porovnání s počtem respondentů není ideální, ovšem je výrazně lepší nežli v sekci sub-testů Vocabulary. Pozitivní stránkou je také průřezová konsistence výsledků hodnot alfa na pilotní sadě testů. V tomto směru se nejmenší a největší hodnota od sebe liší o pouhých 0,1067, což představuje největší soudržnost výstupních hodnot napříč sledovanými skupinami sub-testů.

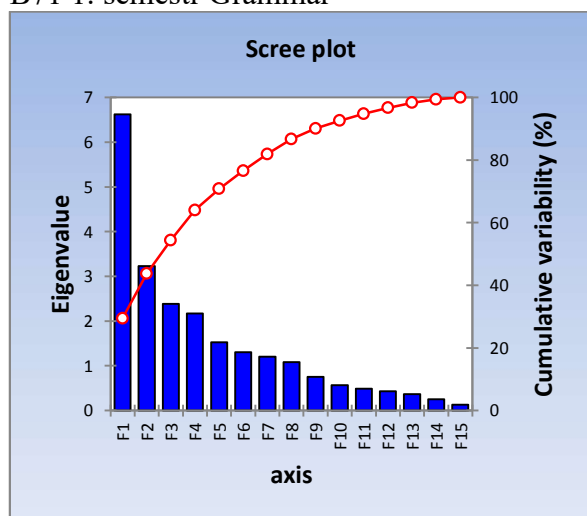
B60 1. semestr Grammar



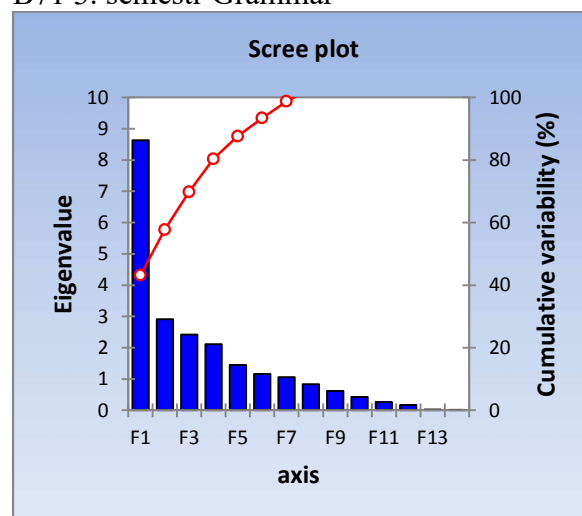
B60 3. semestr Grammar

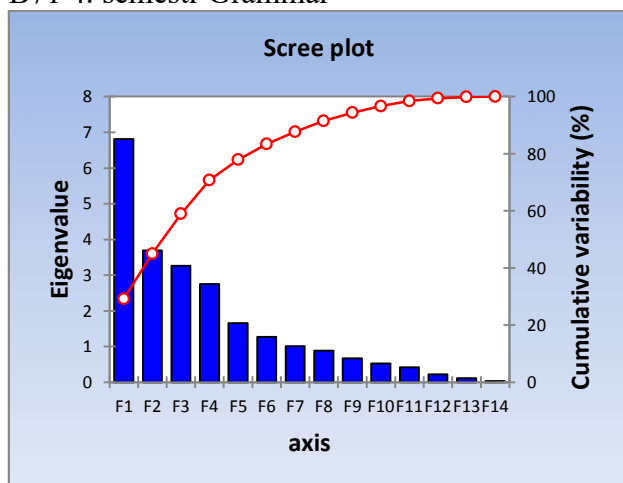


B71 1. semestr Grammar



B71 3. semestr Grammar





Zdroj: Vlastní

Fig. 5: Faktorová zátěž – sub-testy Grammar

Pokud jde o faktory jednotlivých testů, je situace vesměs pozitivní, neboť umožňuje identifikovat sub-testy, které z tohoto pohledu vycházejí lépe, a které jsou naopak problematičtější. Pozitivně v tomto hodnocení vychází sub-test B60 3. semestr, kde celých 12 položek z 20 je soustředěno v dominantním faktoru a zbylých osm rovnoměrně podporuje další faktory. Z opačné škály můžeme vybrat sub-test B71 4. semestr, kde dominantní faktor soustřeďuje 6 položek a zbytek je rozptýlen v dalších faktorech. Konzistence nebo rozptyl položek je graficky demonstrován na srovnávacích grafech na obrázku 5 (Fig. 5).

Lze konstatovat, že celkové hodnocení reliability pro sub-testy Grammar je pozitivní. V těchto sub-testech jsou dosahovány největší hodnoty reliability v porovnání s dalšími skupinami sub-testů. Tuto skutečnost lze vysvětlit relativně velkou praxí v sestavování tohoto typu sub-testů, jakožto i obeznamenost studentů s testovanou oblastí. Pro další smysluplnou interpretaci se jeví jako nezbytné zajištění přiměřeně početného vzorku respondentů ve všech testovaných studijních skupinách, které by dodaly celkovým výsledkům měření na větší relevanci.

Závěr

Je důležité poznamenat, že identifikace pozitivních a problematických příkladů sub-testů z pohledu jejich reliability, která je interpretována v širším kontextu, je důležitým krokem k optimalizaci testovacích procesů v rámci vytváření a doplňování testové banky. Při výzkumu byly odzkoušeny různé postupy (ať již s pozitivním či negativním výsledkem) a z těchto zkušeností je čerpáno při vytvoření postupů tzv. dobré praxe, které jsou aplikovány při následném sestavování dalších testů na katedře.

Literatura

- [1] ALDERSON J. Ch.; CLAPHAM C.; WALL D.: *Language Test Construction and Evaluation*. Cambridge University Press, 1995. ISBN 978-0521478298.
- [2] ALTE: *Manual for Language Test Development and Examining*. [online]. 2011. Available from WWW: <http://www.alte.org/Materials>
- [3] BROWN, J. D.; HUDSON, T.: *Criterion-Referenced Language Testing. The Cambridge Applied Linguistics Series*. Cambridge University Press, 2008. ISB 978-0-521-80628-2.

- [4] CRONBACH, L. J.: Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951, Vol. 16, pp. 297–334. ISSN 0033-3123. DOI: [10.1007/BF02310555](https://doi.org/10.1007/BF02310555)
- [5] FULCHER, G.: *Practical Language Testing*. Routledge, 2013. ISBN-13 978-340-98448-2.
- [6] GREEN, A.: *Exploring Language Assessment and Testing*. Routledge Taylor Francis Group, 2014. ISBN 978-1-415-88962-7.
- [7] HAMBLETON, R. K.; ROVINELLI, R. J.: Assessing the Dimensionality of a Set of Test Items. *Applied Psychological Measurement*. 1986, Vol. 10, Issue 3, pp. 287–302. DOI: [10.1177/014662168601000307](https://doi.org/10.1177/014662168601000307)

TEST CONSTRUCTION AND EMPIRIC TESTING

This article deals with the topic of reliability of language didactic tests in the framework of language testing. Gained data serve as a means for interpretation and moderation of testing at the language department. Tested students are 1st and 2nd year students of the Faculty of Security Management and the Faculty of Security Management and Security and Legal Studies of the Police Academy of the Czech Republic. For determining the reliability the measurable indicator index of inner consistence with subdivision into scales was set, these scales were used for setting the criteria for the evaluation of the reliability of sub-tests. The reliability of sub-tests is interpreted in a wider contest and is an important step to the optimization of the testing process of creating a test bank for the language department.

INTERPRETATION DER UNTERSUCHUNGSFRAGEN DER RELIABILITÄT VON SPRACHTESTS

Dieser Artikel befasst sich mit der Reliabilität von didaktischen Sprachtests im Rahmen von Sprachprüfungen. Die gewonnenen Daten dienen zur Interpretation und der Moderation von Testaufgaben an den Lehrstühlen für Fremdsprachen. Dem Test unterzogen sich Studenten des ersten und zweiten Studienjahres folgender Institutionen: Fakultät für Unmittelbares Management, Fakultät der Akademie der Sicherheitspolizei der Tschechischen Republik in Prag, Fachbereiche Sicherheitsmanagement in der öffentlichen Verwaltung und Sicherheitsrechtliche Studien. Zur Festsetzung der Reliabilität wurde ein messbarer Indikatorindex der inneren Konsistenz mit einer Skaleneinteilung festgelegt, welche zur Festlegung des Kriteriums für die Bewertung der Reliabilität der Subtests verwendet wurde. Die Reliabilität der Subtests wird im weiteren Kontext interpretiert und ist ein wichtiger Schritt zur Optimalisierung der Testprozesse im Rahmen der Bildung und zur Ergänzung einer Testdatenbank für den Lehrstuhl für Fremdsprachen.

INTERPRETACJA PYTAŃ BADAWCZYCH DOTYCZĄCYCH RZETELNOŚCI TESTÓW JĘZYKOWYCH

W artykule omówiono kwestię rzetelności językowych testów dydaktycznych w kontekście egzaminów językowych. Uzyskane dane wykorzystano do interpretacji i moderacji testów w katedrze języków obcych. Badaniami objęto studentów pierwszego i drugiego roku Wydziału Zarządzania Bezpieczeństwem oraz Wydziału Prawa w Zakresie Bezpieczeństwa Akademii Policyjnej Republiki Czeskiej w Pradze, na kierunkach Zarządzanie bezpieczeństwem w administracji publicznej oraz Studia prawnicze z zakresu bezpieczeństwa. W celu określenia rzetelności określono mierzalny wskaźnik spójności wewnętrznej z podziałem na skale, służące określeniu kryteriów do oceny rzetelności podtestów. Rzetelność podtestów jest interpretowana w szerszym kontekście i stanowi ważny krok w kierunku optymalizacji procesów testowania w ramach tworzenia i uzupełniania bazy testów dla katedry języków obcych.