

POLOŽKOVÁ ANALÝZA TESTU Z NĚMECKÉHO JAZYKA POMOCÍ PROGRAMU ITEMAN

Eva Hradilová
* Tereza Krupková

Univerzita obrany Brno
Centrum jazykové přípravy
Kounicova 65, 622 10 Brno, Česká republika
eva.hradilova@unob.cz
* tereza.krupkova@unob.cz

Abstrakt

Příspěvek se zabývá problematikou testování na Centru jazykové přípravy Univerzity obrany v Brně (CJP UO). Konkrétně jde o analýzu a vyhodnocení testu NJ podle normy NATO STANAG 6001 SLP 1 (A1-A2 dle SERR) pomocí statistického softwaru ITEMAN. Cílem výzkumu bylo zjistit, nakolik je daný test validní a reliabilní. Nejprve je charakterizován program ITEMAN, dále je popsán průběh přípravných prací. Největší část článku je věnována položkové analýze a rozboru špatných testových položek. Na závěr je navrženo, jak postupovat při korektuře některých nevhodných položek.

ÚVOD

Příspěvek se zabývá problematikou testování na Centru jazykové přípravy Univerzity obrany v Brně (CJP UO). Konkrétně jde o analýzu a vyhodnocení testu NJ podle normy NATO STANAG 6001 SLP 1 (A1-A2 dle SERR) pomocí statistického softwaru ITEMAN. Tento test bude využíván na Oddělení ostatních jazyků k přezkoušení civilních studentů 1. ročníku Bc. Na základě tohoto ověření znalostí bude studentům udělen zápočet za jeden semestr, pokud dosáhnou alespoň 65 % úspěšnosti. Cílem výzkumu bylo proto zjistit, nakolik je daný test validní a reliabilní, popř. ho na základě provedené položkové analýzy náležitě upravit.

1 Program ITEMAN a přípravné práce

1.1 Program ITEMAN

Program ITEMAN je statistický software americké firmy Assessment Systems Corporation určený pro analýzu jednotlivých testových položek typu multiple choice i chování testu jako celku [4].

Výsledkem analýzy jsou různé číselné údaje, na jejichž základě je možné vyhodnotit, zda jsou jednotlivé úlohy validní a reliabilní či naopak špatné a nespolehlivé. Dále můžeme porovnat úspěšnost lepší a horší třetiny studentů u jednotlivých položek, zjistit dosažené minimum a maximum bodů v rámci celého testu, průměrný výsledek a medián, kterého studenti při tomto přezkoušení dosáhli [3].

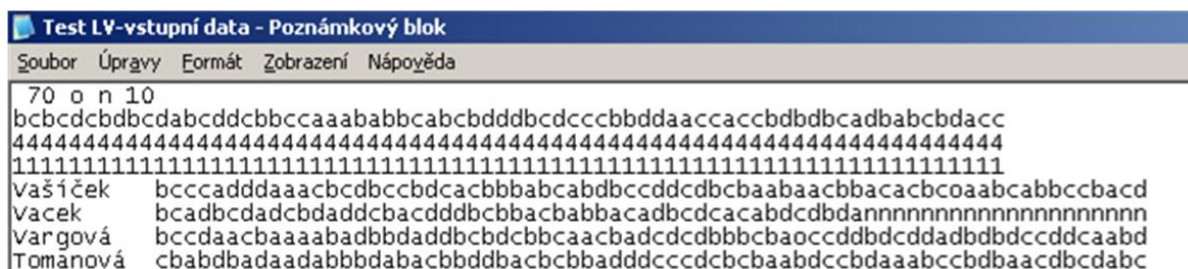
Při souhrnném vyhodnocení celého testu je důležitým údajem koeficient Alpha, který udává, nakolik je test jako celek reliabilní [3].

1.2 Přípravné práce

Námi vytvořený test byl nejprve zmoderován kolegy němčináři. Následně byl pretestován na skupinách studentů, jejichž jazyková úroveň odpovídá SLP 1. Test sestával ze dvou částí. V první části byla testována dovednost poslech s porozuměním, ve druhé potom čtení s porozuměním, které zahrnuje mj. i položky testující gramatické jevy a slovní zásobu.

Vzhledem k počtu testovaných studentů – poslech 71 a čtení 63 – lze považovat výsledná čísla analýzy za dostatečně reprezentativní.

Data z testů byla ručně vkládána do poznámkového bloku, což bylo s ohledem na množství položek – poslech 30 a čtení 70 – a počet studentů časově značně náročné.



Zdroj: vlastní

Fig. 1: Poznámkový blok

Nyní vysvětlíme, co znamenají jednotlivé údaje na Obr. 1 (Fig. 1). Na prvním řádku najdeme počet položek v testu, písmeno „o“ značící vynechanou odpověď, písmeno „n“ značící nestihnutou odpověď a počet znaků pro napsání identifikačních údajů studentů (např. příjmení). Do druhého řádku se zaznamenává klíč správných odpovědí. Ve třetím řádku najdeme počet nabízených možností a ve čtvrtém řádku se uvádí počet správných odpovědí. Od pátého řádku se pak vkládají identifikační údaje studentů a jejich odpovědi u jednotlivých položek [1].

Tento soubor je následně vložen do programu ITEMAN, který provede první vyhodnocení. Na základě tohoto musí být provedena kontrola údajů, aby se vyloučily překlepy nebo chybné součty bodů při opravě, což by zkreslilo celkové výsledky. Takto zrevidovaný soubor je podruhé vyhodnocen programem ITEMAN a pak následuje vlastní položková analýza.

2 Položková analýza

V našem případě jsme analyzovaly celkem 100 položek, z nichž každá má po zpracování programem ITEMAN následující podobu viz Obr. 2 (Fig. 2).

Item Statistics					Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
22	1-22	.62	.59	.53	A	.62	.35	.94	.53 *
					B	.27	.53	.06	-.48
					C	.03	.00	.00	.01
					D	.08	.12	.00	-.16
					Other	.00	.00	.00	.00

Zdroj: vlastní

Fig. 2: Položková analýza-jednotliví ukazatelé

Při analýze se pracuje s jednotlivými ukazateli [3]:

- Seq. No. – pořadové číslo položky,
- Scale-Item – pořadové číslo položky v subtestu,
- Prop. Correct – podíl/procento správných odpovědí (0 až 1), ideální 0,3-0,8,

- Disc. Index – citlivost položky/rozdíl mezi úspěšností lepší a horší třetiny (-1 až +1), přijatelné od 0,3,
- Point Biser. – korelace úspěšnosti u položky s celkovým skórem testovaného (-1 až +1), přijatelné od 0,3,
- Alt. – distraktory,
- Prop. Total – podíl/procento odpovědí u jednotlivých distraktorů celkem (0 až 1),
- Endorsing Low – podíl/procento odpovědí u jednotlivých distraktorů v horší třetině (0 až 1),
- Endorsing High – podíl/procento odpovědí u jednotlivých distraktorů v lepší třetině (0 až 1),
- Key – označení správné odpovědi.

2.1 Dobré položky

Příkladem dobré položky je např. číslo 22 v části čtení s porozuměním, její statistické vyhodnocení viz Obr. 2 (Fig. 2). Obtížnost úlohy je přiměřená – správně ji zodpovědělo 62 % studentů, ideálně by se tento ukazatel měl pohybovat v rozmezí 30 % – 80 %. Položka je dostatečně citlivá, protože dobře rozlišuje horší studenty od lepších – správně odpovědělo 94 % studentů z lepší třetiny, ale jen 35 % z horší třetiny, rozdíl je tedy 59 %, přičemž přijatelná hodnota začíná od 30 %. Posledním důležitým ukazatelem kvality položky je Point Biser., tedy korelace úspěšnosti u dané úlohy s celkovým skórem testovaného. Pokud nabývá hodnoty 0,3 a více (čím blíže k 1,0, tím lépe), dokáže dobře předpovědět celkový výsledek studenta v testu. V případě této konkrétní položky se jedná o hodnotu 0,53.

Takto vypadala daná úloha v testu:

22. Ich bin ernst krank, mein Arzt hat mir Arznei -----.

- a) verschrieben
- b) geschrieben
- c) nachgeschrieben
- d) zugeschrieben

2.2 Špatné položky

Na základě položkové analýzy jsme rozdělily nevyhovující položky do čtyř následujících kategorií:

- příliš snadné,
- příliš obtížné,
- nerozlišující mezi horšími a lepšími studenty,
- chybně předpovídající výsledek studenta v celém testu.

2.2.1 Příliš snadné položky

V rámci celého testu (100 úloh) se ukázalo, že příliš snadných položek bylo pět. Na jednom příkladu budeme demonstrovat, jak tento typ špatné položky vyhodnotil program ITEMAN – Obr. 3 (Fig. 3).

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing		Point Biser.	Key
							Low	High		
39	1-39	.98	.06	.30	A	.02	.06	.00	-.30	
					B	.00	.00	.00		
					C	.00	.00	.00		
					D	.98	.94	1.00	.30	*
					Other	.00	.00	.00		

Zdroj: vlastní

Fig. 3: Položková analýza-ukázka příliš snadné položky

Jednalo se o číslo 39 v části čtení s porozuměním. Úloha je příliš snadná – správně ji zodpovědělo 98 % studentů, a tudíž není dostatečně citlivá, neboť rozdíl mezi horší a lepší třetinou studentů činí pouze 6 %. Poslední důležitý ukazatel kvality položky Point Biser. vykazuje hraniční přípustnou hodnotu 0,3.

Zadání v testu vypadalo následovně:

39. Die Schwester meiner Mutter ist meine:

- a) Nichte
- b) Oma
- c) Frau
- d) Tante

2.2.2 Příliš obtížné položky

V rámci celého testu (100 úloh) se ukázalo, že příliš obtížné položky byly pouze tři. Na následujícím příkladu ukážeme, jak tento typ špatné položky vyhodnotil program ITEMAN – Obr. 4 (Fig. 4).

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing		Point Biser.	Key
							Low	High		
59	1-59	.25	.61	.52	A	.17	.18	.17	.06	
					B	.25	.12	.11	-.01	
					C	.25	.00	.61	.52	*
					D	.22	.47	.11	-.37	
					Other	.10	.00	.00	-.31	

Zdroj: vlastní

Fig. 4: Položková analýza-ukázka příliš obtížné položky

Jednalo se o číslo 59 v části čtení s porozuměním. Položka je příliš obtížná – správně ji zodpovědělo jen 25 % studentů. Jinak je ale dostatečně citlivá – rozdíl mezi horší a lepší třetinou studentů je 61 % – i dobře předpovídá výsledek studentů v celém testu, neboť poslední důležitý ukazatel Point Biser. vykazuje solidní hodnotu 0,52.

Takto vypadala daná položka v testu:

59. Neue Erfahrung

Die Schule war anders als in Deutschland. Ich hatte jeden Tag acht bis neun Stunden. Am Anfang habe ich natürlich sehr schlecht Spanisch gesprochen. Die Schüler in meiner Klasse

waren toll! Im Februar haben wir Theater gespielt, es war chaotisch, stressig, aber wir haben so viel gelacht.

Worüber spricht die Person?

- a) Über ihren Feriensprachkurs.
- b) Über ihre Klassenfahrt nach Spanien.
- c) Über ihren Austauschaufenthalt.
- d) Über ihren Theaterkurs.

2.2.3 Položky nerozlišující mezi horšími a lepšími studenty

V rámci celého testu (100 položek) se ukázalo, že tyto úlohy byly čtyři. Na dalším příkladu budeme demonstrovat, jak tento typ špatné položky vyhodnotil program ITEMAN – Obr. 5 (Fig. 5).

Item Statistics					Alternative Statistics					
Seq. No.	Scale -Item	Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low High		Point Biser.	Key
11	1-11	.27	-.13	.07	A	.17	.24	.11	-.16	
					B	.54	.29	.61	.14	?
					C	.00	.00	.00		
					D	.27	.41	.28	.07	*
					Other	.02	.00	.00	-.30	

CHECK THE KEY
d was specified, b works better

Zdroj: vlastní

Fig. 5: Položková analýza-ukázka položky nerozlišující mezi horšími a lepšími studenty

Jednalo se o číslo 11 v části čtení s porozuměním. Položka je poměrně obtížná – správně ji zodpovědělo jen 27 % studentů. Další ukazatele jsou ale ještě mnohem horší – rozdíl mezi horší a lepší třetinou studentů nabývá záporných hodnot (-13 %), protože tento úkol správně vyřešilo 41 % studentů z horší třetiny, ale jen 28 % studentů z lepší třetiny. Většina dobrých studentů si naopak zvolila možnost B, proto je u tohoto typu špatných položek nutné ověřit správnost klíče. Tato úloha navíc ne zcela dobře předpovídá výsledek studentů v celém testu, neboť poslední ukazatel Point Biser. je pouze 0,07.

Takto vypadala daná položka v testu:

11. Das Fernsehen überträgt heute ein Fußballspiel. Ich -----.

- a) sehe es sich an.
- b) sehe es mich an.
- c) ansehe es mir.
- d) sehe es mir an.

2.2.4 Položky chybně předpovídající výsledek studenta v celém testu

V rámci celého testu (100 úloh) se ukázalo, že položky chybně předpovídající výsledek studenta v celém testu byly tři. Na následujícím příkladu ukážeme, jak tento typ špatné položky vyhodnotil program ITEMAN – Obr. 6 (Fig. 6).

Seq. No.	Item Statistics				Alternative Statistics					
	Scale -Item	Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser.	Key
15	1-15	.24	-.13	-.11	A	.40	.47	.33	-.14	
					B	.30	.12	.44	.34	?
					C	.02	.06	.00	-.13	
					D	.24	.29	.17	-.11	*
					Other	.05	.00	.00	-.12	

Zdroj: vlastní

Fig. 6: Položková analýza-ukázka položky chybně předpovídající výsledek studenta v celém testu

Jednalo se o číslo 15 v části čtení s porozuměním. Položka je poměrně obtížná – správně ji zodpovědělo jen 24 % studentů. Další dva ukazatele jsou však ještě podstatně horší – rozdíl mezi horší a lepší třetinou studentů nabývá podobně jako v předcházejícím příkladu 2.2.3 záporných hodnot (-13 %), protože tento úkol správně vyřešilo 29 % studentů z horší třetiny, ale jen 17 % studentů z lepší třetiny. Většina dobrých studentů si naopak zvolila možnost B, proto je nutné ověřit správnost klíče. Tato položka navíc vůbec nepředpovídá správně výsledek studentů v celém testu, neboť poslední ukazatel Point Biser. je rovněž záporný (-0,11).

Úloha v testu vypadala následovně:

15. Gehst du auch hin? – Ich kann nicht hingehen, denn -----.

- habe ich keine Zeit.
- ich keine Zeit habe.
- keine Zeit ich habe.
- ich habe keine Zeit.

3 Aplikace výsledků položkové analýzy

Na základě položkové analýzy se nám podařilo rozdělit položky na dobré a špatné. Aby byl test co nejkvalitnější, je žádoucí se špatnými úlohami dále pracovat. Pro jednotlivé typy navrhuje následující možnosti postupu:

- příliš snadné** – umístit na začátek/doprostřed testu, popř. ztížit,
- příliš obtížné** – zjednodušit (zkrátit distraktory, přeformulovat) nebo použít v testu vyšší úrovně,
- nerozlišující mezi horšími a lepšími studenty** – vyřadit z testu, přeformulovat, vyhnout se výjimkám,
- chybně předpovídající výsledek studenta v celém testu** – vyřadit z testu, přeformulovat, vyhnout se výjimkám.

A nyní přejdeme k ukázce toho, jak jsme při aplikaci výsledků postupovaly u výše zmíněných čtyř konkrétních příkladů špatných položek.

Ad a) U položky číslo 39 jsme se rozhodly ji ztížit jednak přeformulováním zadání, jednak volbou obtížnějších distraktorů.

Návrh nového znění úkolu:

39. Die Tochter meiner Schwester ist meine:

- a) Nichte
- b) Cousine
- c) Enkelin
- d) Schwiegertochter

Ad b) Položku číslo 59, která je až na vysokou obtížnost v jiných kritériích validní, jsme ponechaly v původní podobě a využijeme ji při tvorbě testu na vyšší úroveň.

Ad c) U úlohy číslo 11 jsme zachovaly typ gramatického jevu, ale vybraly jsme jiné sloveso, jehož rekcí si studenti snáze zapamatují.

Návrh nové podoby zadání:

11. Wir kennen uns noch nicht. _____

- a) Ich stelle sich vor.
- b) Ich stelle mich vor.
- c) Ich stelle mir vor.
- d) Ich stelle dir vor.

Ad d) Zadání číslo 15 jsme přeformulovaly, přičemž jsme opět zachovaly typ gramatického jevu, ale použily jsme jinou spojku souřadící, kterou studenti nezaměňují se synonymní spojkou podřadící.

Návrh nové položky:

15. Gehst du auch hin? – Nein, aber -----.

- a) es tut mir sehr Leid.
- b) tut es mir sehr Leid.
- c) es mir sehr Leid tut.
- d) tut mir es sehr Leid.

Závěr

Vzhledem k celkovým výsledkům vyhodnocení našeho testu programem ITEMAN, kdy hodnota Alpha u čtení s porozuměním dosahuje 0,890 a u poslechu s porozuměním 0,839, jej lze označit za vysoce validní a reliabilní. Obecně lze říci, že pokud je ukazatel Alpha vyšší než 0,6, je test přijatelný, a čím více se blíží 1,0, tím je kvalitnější.

Proto není nezbytně nutné test přepracovat, nicméně na základě položkové analýzy bylo vytipováno 15 úloh, se kterými by bylo vhodné ještě dále pracovat. Některé budou vypuštěny, další přeformulovány nebo přemístěny v rámci stávajícího testu, popř. použity v testech jiné úrovně. Tímto způsobem upravený test bude poté použit na CJP UO.

Práce s programem ITEMAN se nám osvědčila. Uvědomily jsme si například, že některé položky, které nám při tvorbě přišly naprosto jednoznačné, se ve skutečnosti projeví jako zavádějící. Vzhledem k faktu, že nás čeká vytváření testů na další úrovně, počítáme i v budoucnu s využitím tohoto softwaru a nově nabytých zkušeností.

Literatura

- [1] BYČKOVSKÝ, P., MARKOVÁ, M., 2003. Využití software ITEMAN k položkové analýze a analýze výsledků testů. In *Sociální a kulturní souvislosti výchovy a vzdělávání: 11. výroční mezinárodní konference ČAPV: Sborník referátů* [CD-ROM]. Brno: Masarykova univerzita, Pedagogická fakulta, 2003.

- [2] Handbuch zur Entwicklung und Durchführung von Sprachtests. ALTE. Frankfurt am Main: telc GmbH, 2012. ISBN: 978-3-86375-093-0.
- [3] User's Manual for the ITEMAN Conventional Item Analysis Program. Assessment Systems Corporation.
- [4] *Iteman 4 - Test and item analysis software with classical test theory* [online]. 2010 [cit. 2013-09-28]. Dostupné z: <<http://assess.com/xcart/product.php?productid=417>>.

THE ITEM ANALYSIS OF GERMAN LANGUAGE TESTS THROUGH THE ITEMAN APPLICATION

The paper deals with the testing at the Language Training Centre of the University of Defense in Brno (CJP UO). Specifically, the analysis and the German tests evaluation following the NATO STANAG 6001 SLP 1 (A1-A2 CEFR) standards through the statistical software ITEMAN is considered. The aim of the research was to find out whether the test is valid and reliable. Firstly, the application ITEMAN is described, further on the paper carries on with the process of preliminary work. The largest part of the article is devoted to the analysis and item analysis of the bad test items. Finally, the procedures for some inappropriate items proofreading are proposed.

ANALYSE VON DEUSCHTESTS MIT HILFE DES PROGRAMMS ITEMAN

Der Beitrag befasst sich mit der Problematik des Testens am Zentrum für Fremdsprachenausbildung an der Verteidigungsuniversität in Brünn (CJP UO). Konkret handelt es sich um die Analyse und Auswertung eines Deutschtestes gemäß des NATO-Standards STANAG 6001 Leistungsstufe 1 (A1-A2 gemäß GeR) mittels der statistischen Software ITEMAN. Das Forschungsziel war festzustellen, inwieweit der Test valide und reliabel ist. Anfangs wird das Computerprogramm ITEMAN vorgestellt, weiter wird das Vorbereitungsverfahren beschrieben. Der wesentliche Teil des Artikels betrifft die Aufgabenanalyse und Erörterung schlechter Testaufgaben. Zum Schluss wird vorgeschlagen, wie einige von diesen nicht geeigneten Testaufgaben zu verbessern sind.

ANALIZA ZADAŃ TESTOWYCH JĘZYKA NIEMIECKIEGO PRZY POMOCY PROGRAMU ITEMAN

W artykule przedstawiono sposób przeprowadzania testów w Centrum Przygotowania Językowego Uniwersytetu Obrony w Brnie. Konkretnie opracowanie dotyczy analizy i oceny testu języka niemieckiego według normy NATO STANAG 6001 SLP 1 (A1-A2 wg poziomów biegłości językowej CEFR) przeprowadzonej przy pomocy oprogramowania statystycznego ITEMAN. Celem przeprowadzonych badań było stwierdzenie, na ile dany test jest stosowny i rzetelny. W pierwszej kolejności scharakteryzowano program ITEMAN, następnie opisano przebieg prac przygotowawczych. Największa część artykułu dotyczy analizy poszczególnych zadań testowych oraz zbadania zadań nieprawidłowych. Na zakończenie przedstawiono propozycję korekty niektórych niewłaściwych zadań.