

**Masarykova univerzita**

# **Adaptivní test COMPACT**

**Tamara Váňová  
Jiří Procházka  
Denisa Denglerová**

**Brno 2012**

Publikace vznikla v rámci projektu CZ.1.07/2.2.00/07.0442 Kompetence v jazykovém vzdělávání na Masarykově univerzitě – COMPACT. Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Odborná recenze: prof. PaedDr. Iva Stuchlíková, CSc.

©Masarykova univerzita, 2012

ISBN 978-80-210-5742-5

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Počítačové adaptivní testování</b>	<b>9</b>
2.1	Počátky testování, vznik a vývoj adaptivních testů . . . . .	9
2.2	Přínosy počítačového adaptivního testování . . . . .	17
2.3	Průběh počítačového adaptivního testování . . . . .	19
2.3.1	Položková banka . . . . .	19
2.3.2	Začátek testování . . . . .	20
2.3.3	Vlastní testování . . . . .	21
2.3.4	Ukončení testu . . . . .	21
2.4	Matematické teorie stojící v pozadí CAT . . . . .	22
2.4.1	Základní principy teorie odpovědi na položku . . . . .	22
2.5	IRT modely . . . . .	24
2.6	Informační funkce položky v kontextu IRT . . . . .	26
		<b>3</b>

## OBSAH

---

<b>3</b>	<b>Vznik a dosavadní vývoj adaptivního testu COMPACT</b>	<b>31</b>
3.1	Východiska . . . . .	31
3.2	Vývoj testu . . . . .	37
3.3	Využívání testu . . . . .	43
3.4	Podzimní semestr 2011 . . . . .	46
3.4.1	Je požadavek vstupní úrovně B1 oprávněný? .	47
3.5	Technické parametry . . . . .	48
<b>4</b>	<b>Testovací prostředí COMPACT</b>	<b>51</b>
4.1	Jak se vybírají otázky do rozřazovacího testu . . . . .	51
4.2	Průvodce testem COMPACT . . . . .	53
4.3	Průchod testem v roli studenta . . . . .	54
4.4	Jak se systémem pracovat v dalších rolích . . . . .	64
4.4.1	Stručný popis nových rolí . . . . .	64
4.4.2	Pohled učitele . . . . .	66
4.4.3	Správa položek . . . . .	67
4.4.4	Statistické informace o položkách . . . . .	77
4.4.5	Nastavení vlastních testů . . . . .	78
4.4.6	Jak pretestovat . . . . .	86
4.4.7	Administrace výsledků a jejich převod do IS .	87
<b>5</b>	<b>Závěr</b>	<b>93</b>
<b>6</b>	<b>Literatura</b>	<b>95</b>

# Kapitola 1

## Úvod

Sebelepší počítačem podporovaný test zatím není schopen dostatečně komplexně otestovat schopnosti studenta funkčně používat cizí jazyk. Přesto jsme s díky přijali výzvu vytvořit adaptivní rozřazovací test z anglického jazyka s diagnostickými prvky, který má, coby jeden z výstupů evropského projektu COMPACT (Kompetence v jazykovém vzdělávání na Masarykově univerzitě), sloužit vyučujícím i studentům nejen jako hotový produkt, ale i jako nástroj pro další rozvoj standardizovaného, potažmo adaptivního testování cizích jazyků na Masarykově univerzitě (MU).

Na jedné straně je jeho záměrem šetřit čas a energii učitelům, aby nemuseli testovat jednotlivé skupiny studentů v prvních týdnech semestru, ale zároveň jim chce poskytovat jistotu, že se v seminárních skupinách nebudou vyskytovat studenti s nedostatečnou vstupní jazykovou úrovní, protože bez přiměřené vstupní prakticky nelze ve většině běžných případů dosáhnout plánované úrovně výstupní, která je pro celou MU vyhlášena jednotně.

Studenty, na druhé straně, má tento nástroj včas (ideálně v prv-

ním semestru studia hned po zápisu do prvního ročníku, ale v praxi k tomu dochází často až před registrací do některého z předmětů odborné angličtiny v semestru podle volby studenta) upozornit, že pokud nedosahují požadované úrovně pro úspěšné studium odborného cizího jazyka ve zvoleném studijním oboru na MU, je načase s tím urychleně něco dělat – navštěvovat jazykovou školu, doučit se jazyk soukromě, odstranit nedostatky formou samostudia apod. Navíc díky získaným diagnostickým informacím se studenti mohou při svém dalším studiu zaměřit na konkrétní gramatické a lexikální okruhy, jež test identifikoval jako nedostatečně zvládnuté s ohledem na celkovou cizojazyčnou úroveň. Součástí informací po absolvování testu je i doporučení, která z běžně dostupných učebnic samostudijního typu je vhodná pro jejich úroveň znalostí/neznalostí a do kterého z remediálních jazykových předmětů nabízených studentům MU se mohou zaregistrovat.

Tato publikace si klade za cíl být především průvodcem pro učitele Centra jazykového vzdělávání na Masarykově univerzitě (CJV MU), jak s tímto prostředím pracovat nejen jako s hotovým diagnostickým nástrojem, ale jak ho používat pro postupné vytváření vlastních testů ať už lineárních, či adaptivních. Po vytvoření dostatečně rozsáhlé testové banky s náležitě kalibrovanými položkami ho mohou používat i pro klasifikaci výsledků studentů jednotlivých pracovišť CJV MU.

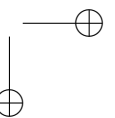
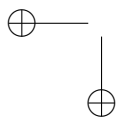
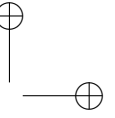
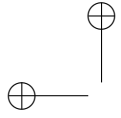
Vzhledem k tomu, že v českém kontextu příliš mnoho informací o adaptivním testování zatím není, první kapitola se tomuto způsobu hodnocení věnuje z obecného hlediska. Nemůže v ní tudíž chybět ani jeho historie a nástin teoretických principů. Druhá kapitola popisuje důvody vzniku tohoto nástroje, jeho postupný vývoj a praktické využívání na vybraných fakultách MU, které se do této části projektu zapojily. Součástí této kapitoly je také stručné shrnutí technických informací o konkrétním softwarovém prostředí. Třetí kapitola obsa-

---

huje ilustrační příklady toho, co studenta při průchodu testem čeká, a pak hlavně návod, jak s nástrojem mohou učitelé pracovat a dále ho rozvíjet.

Publikace nastoluje i jednu velmi důležitou otázku: Kde přesně leží hranice mezi jednotlivými úrovněmi podle Společného referenčního rámce pro jazyky (SERRJ)? Na ni zatím nepřinesla jednoznačnou odpověď ani práce na tomto projektu, ani spolupráce na jiném evropském projektu, který pod názvem English Profile Project tyto hranice mapuje na mezinárodní úrovni pod vedením Cambridge University. Je pochopitelné a vyplývá to i z popisu jednotlivých úrovní SERRJ, že žádný lexikálně-gramatický test není izolovaně sám o sobě schopen bezpečně a spolehlivě rozlišovat mezi úrovněmi B2 a C1/C2. K tomu by bylo zapotřebí testovat studenty i v celé řadě jiných jazykových dovedností, jako je například samostatný ústní a písemný projev nebo schopnost zpracovávat informace z cizojazyčného textu či poslechu. Definovat přesnou hranici mezi B2 a vyššími úrovněmi však ani nebylo naším cílem. Tato část projektu měla za úkol vytvořit takový nástroj, který by byl schopen odlišit studenty s úrovní A1/A2 od úrovně B1, neboť ta je deklarovanou vstupní úrovní pro studium odborné angličtiny na fakultách MU.

Přesto však, díky tomu, že tento nástroj není statický, ale lze ho dynamicky dále rozvíjet, není vyloučené, aby se v průběhu času a za pomoci dalších nástrojů počítačového testování schopných simulovat mezilidskou komunikaci, měřit kvalitu výslovnosti, posuzovat například tzv. metodou data mining úroveň testanta podle aktivního používání pokročilé slovní zásoby apod., nemohlo jádro této části projektu COMPACT integrovat do pokročilých technologií ICT a přispět ke standardizaci úrovně cizojazyčných znalostí studentů napříč fakultami MU.





## Kapitola 2

# Počítačové adaptivní testování

Tato kapitola se věnuje historii a teoretickým principům včetně matematických teorií, které stojí v pozadí počítačového adaptivního testování.

### 2.1 Počátky testování, vznik a vývoj adaptivních testů

Zřejmě první zmínky o aktivitách, které bychom mohli chápat v dnešních pojmech jako testování, spadají do období starověké Číny. Již kolem roku 2200 př. n. l. se objevují ústní a výkonové zkoušky při výběru úředníků do státní správy (Svoboda, 2010). Oblasti testování zahrnovaly znalosti aritmetiky, zákonodárství, dovednosti lukostřelby a jízdy na koni a další.

## Počítačové adaptivní testování

---

Vědecký přístup k testování se však vyvíjí mnohem později, až v průběhu 19. století. V první polovině 19. století sílily tři nezávislé tendence, které ovlivnily vývoj testování. Jednalo se jednak o studium osob s psychickými abnormalitami reprezentované především francouzskou psychiatrickou komunitou, dále o experimentální výzkum chování zdravých dospělých reprezentovaný německou školou psychofyziky a nakonec o matematické koncepcí normálního rozložení pravděpodobnosti (Urbánek, Denglerová, Širůček, 2011). Zjednodušeně můžeme říci, že z těchto nastíněných koncepcí se vyvinula klasická testová teorie, jejíž princip je založen na předložení určitého testu velké skupině lidí a na základě jejich odpovědí (na stejné otázky) jsou statisticky stanoveny normy vystihující danou populaci.

Dalším posunem v testování je přikročení k adaptivním testům. Jedná se o takový způsob testování, kdy je každému probandovi předložena jiná množina takových položek, aby efektivně zachytila jeho schopnosti (princip bude objasněn později). K výraznému rozvoji adaptivního testování dochází právě v dnešní době. Dva základní pilíře, na nichž stojí, jsou vhodná matematická teorie (např. teorie odpovědi na položku, teorie vědomostního prostoru, teorie měření rozhodování apod.), jež poskytuje teoretická východiska a matematický aparát, a potenciál výpočetní techniky, s jehož využitím jsme schopni v krátkém čase zpracovat a také uchovávat ohromné množství dat. Přes výše zmíněné však adaptivní testování není výdobytkem třetího tisíciletí, v mnoha starších testech se objevují jeho náznaky.

Počátky adaptivního testování se objevily již více než před sto lety v roce 1905 v rámci testu francouzského psychologa Alfreda Bineta (1857-1911). Binet se celý život profesionálně zabýval výzkumem jedinců odlišných od normy a zvláště se zaměřil na zkoumání mentálně retardovaných dětí. Jeho cílem bylo vytvořit nástroj, který by hned na počátku (či před započítáním) školní docházky dokázal identifikovat

## 2.1 Počátky testování, vznik a vývoj adaptivních testů

---

děti, které budou v běžné základní škole selhávat a kterým by měla být vzhledem k jejich mentálnímu deficitu poskytnuta speciální péče. V roce 1905 spolu se svým kolegou Theodorem Simonem vytvořili první škálu, která měřila inteligenci dětí, v letech 1908 a 1911 došlo ještě k revizím těchto škál (Urbánek, Denglerová, Širůček, 2011).

Nyní si objasníme podrobněji princip tohoto testu, neboť vhodně vystihuje princip adaptivních testů obecně. Binet-Simonův test se skládal z množiny úloh seřazených dle chronologického věku, pokrýval rozsah 3-11 let. Úlohy byly přiřazeny do dané věkové úrovně, pokud je správně zodpovědělo 50% dětí daného věku (Filípková, Byčkovský, 2008). Vycvičený psycholog obvykle začal s úlohami věkové úrovně, jež odpovídala kalendářnímu věku dítěte, a pokud je dítě zodpovědělo správně, pokračoval s úlohami z vývojově vyšší úrovně. Pokud je však dítě nedokázalo správně zodpovědět, další otázky, které dostalo, byly jednodušší, odpovídající nižší vývojové úrovni.

Takto testování pokračovalo, dokud nedošlo ke stanovení dolní věkové meze (věková úroveň, na níž dítě správně zodpoví všech 10 otázek) a horní věkové meze (první věková úroveň, na níž dítě neodpoví na žádnou otázku správně) (Filípková, Byčkovský, 2008).

Stejná myšlenka je dodnes uchována ve velmi rozšířených testech Wechslerových. Respondent začíná odpovídat na položky na nějaké úrovni a po několika neúspěších je testování adaptivně zastaveno (nejsou mu rigidně předkládány všechny další položky daného subtestu, protože postupně roste pravděpodobnost neúspěšných odpovědí). K administraci takto adaptivních testů je zapotřebí na jednoho testovaného jeden testující, který aktivně (podle předem určených pravidel) vybere další vhodnou otázku, nelze je tedy realizovat skupinově.

Další pokrok ve vývoji adaptivního testování je spojován s vývojem výpočetní techniky. Je nutno poznamenat, že k hlavnímu rozvoji po-

## Počítačové adaptivní testování

---

čítačově adaptivních testů docházelo a dochází v USA, v Evropě se tento princip testování prosazuje pomaleji. První jednoduché počítačové adaptivní testy se objevily v 50. letech 20. století. Dalších 20 let se jednalo o tzv. poloadaptivní testy, kdy byly nějakým algoritmem dopředu stanoveny různé možnosti cest a smyček, jak je možno odpovídat, v podstatě však nepokrývaly všechny teoreticky možné kombinace odpovědí na položky. Zjednodušeně si můžeme tyto testy představit jako definované mnoha pravidly typu „pokud odpověď na otázku 1 zní ano, přejdi k otázce číslo 13“ apod.

Zlom nastává v roce 1973, kdy David J. Weiss (nyní profesor na Minnesotské univerzitě) navrhl počítačovou variantu Binet-Simonova testu, kterou nazval stratifikovaný neboli stradaptivní test. „. . . Weissův test používal stejnou strukturu banky úloh jako Binetův test, úlohy byly uspořádány po deseti do věkových mentálních úrovní dle obtížnosti (tzv. strata = vrstva). Podobně jako v Binetově testu používá stratifikovaný test proměnlivou startovací úroveň, a tím dovoluje začít na jakékoli úrovni obtížnosti přiměřené každému testovanému.

Stratifikovaný test se od Binetova liší v tom, že je zadána vždy pouze jedna úloha a skórována. Další úloha je zadána na základě předchozí odpovědi. Pokud je odpověď správná, je testovanému zadána úloha z následující těžší vrstvy. Když testovaný odpoví chybně, bude mu zadána úloha z nejbližší nižší vrstvy. Tento proces pokračuje tak dlouho, dokud není splněno kritérium k ukončení testu. Test je ukončen, když jsou všechny úlohy nebo pět po sobě následujících úloh v určité vrstvě zodpovězeny chybně. . . “ (Filípková, Byčkovský, 2008).

V následujících desetiletích bylo vytvořeno několik dalších počítačově adaptivních verzí původního Binet-Simonova testu a zároveň začaly být do této podoby adaptovány i další metody a také tvořeny další testy z různých oblastí již přímo v počítačově adaptivní formě. Po roce 2000 se tento způsob testování začíná více objevovat i v Evropě.

## 2.1 Počátky testování, vznik a vývoj adaptivních testů

---

Největší boom zažívá počítačové adaptivní testování (dále CAT, z anglického Computerized Adaptive Testing) v oblasti tzv. výkonových testů (tedy u testů, kdy je jasná správná odpověď, v psychologii např. testy inteligence, paměťových schopností apod.) a v oblasti pedagogického testování vědomostí, znalostí či schopností jedince. V oblasti pedagogického testování se nejčastěji můžeme setkat s testy matematických schopností (je relativně jednoduché sestavit množinu matematických úloh z hlediska vzrůstající obtížnosti) a jazykových schopností, obzvláště angličtiny jako druhého jazyka (Společný evropský referenční rámec pro jazyky usnadňuje tvorbu testovacích nástrojů tím, že mimo jiné definuje šest úrovní ovládnutí jazyka, které umožňují měřit pokroky studentů).

Další výrazné uplatnění nachází CAT v e-learningových adaptivních systémech (Albert, Hockemeyer, Wesiak, 2002). Dobře fungující e-learningové adaptivní systémy v sobě musí obsahovat i modul zhodnocení vědomostí, aby bylo možné rozhodnout, čemu daného jedince nadále učit, tzn. musejí obsahovat způsob testování. V některých systémech je toto testování zcela zjevné, v jiných více maskovaného, například ve výukových počítačových hrách.

Aktuálním trendem je snaha o uplatnění CAT také mimo výkonové testy, zejména v oblasti osobnostních testů (Jelínek, Květon, Denglerová, 2006). V těchto situacích při odpovědi na testovou položku samozřejmě nehodnotíme, zda jsou otázky dobře či špatně zodpovězeny, ale používáme formulaci odpovědi v indikovaném směru (Urbánek, Šimeček, 2001). Např. v testu extroverze považujeme za položky zodpovězené v indikovaném směru takové, které svědčí o vysoké úrovni tohoto rysu osobnosti.

## Počítačové adaptivní testování

Název testu či e-learningového nástroje	Zkratka	Autoři	Popis nástroje
Internetový odkaz			
Graduate Management Admission Test	GMAT	ETS (USA)	Test pro potřeby Graduate Management Admission Council
<a href="http://www.review.cz/www.mba.com/TaketheGMAT">http://www.review.cz/www.mba.com/TaketheGMAT</a>			
Graduate Record Examination	GRE	ETS (USA)	Test studijních předpokladů používaný v USA při přijímacím řízení na postgraduální studium
<a href="http://www.ets.org/portal/site/ets">www.ets.org/portal/site/ets</a>			
National Council Licensure Examinations	NCLEX	NCSBN (National Council of State Boards of Nursing; USA)	Test k udělení licence pro zdravotní sestry
<a href="http://www.ncsbn.org/nclex">www.ncsbn.org/nclex</a>			
Armed Services Vocational Aptitude Test Battery	ASVAB	U.S. Department of Defense (USA)	Multiple-ability test battery
<a href="http://www.usmilitary.com/placementtests">www.usmilitary.com/placementtests</a>			

## 2.1 Počátky testování, vznik a vývoj adaptivních testů

Název testu či e-learningového nástroje	Zkratka	Autoři	Popis nástroje
Internetový odkaz			
Adaptive Matrices Test	AMT	Dr. Schuhfried GmbH (Rakousko)	Nonverbální IQ test založený na deduktivním úsudku
<a href="http://www.schuhfried.at/eng/wts/amt">www.schuhfried.at/eng/wts/amt</a>			
CAT of Written English for Spanish Speakers	eCat	CAT research group at the Autonoma University of Madrid	Test z angličtiny pro Španěle, je zadáván on-line
<a href="http://www.iic.uam.es/pdfs/eCatPDF.pdf">www.iic.uam.es/pdfs/eCatPDF.pdf</a>			
Computerized Adaptive Test of English	CATE	English Language and Learning Support of the Information and Learning Resource Services at Middlesex University (Velká Británie)	Test angličtiny pro uchazeče o studium, kteří nemají angličtinu jako rodný jazyk
<a href="http://www.ilrs.mdx.ac.uk/lang">www.ilrs.mdx.ac.uk/lang</a>			
Assessment and Learning in Knowledge Spaces	ALEKS	ALEKS Corporation (USA)	E-learningový systém k výuce matematiky pro všechny stupně pokročilosti – od ZŠ až po specializované univerzitní kurzy
<a href="http://www.aleks.com/">www.aleks.com/</a>			

## Počítačové adaptivní testování

Název testu či e-learningového nástroje	Zkratka	Autoři	Popis nástroje
Internetový odkaz			
Responsive Open Learning Environments	ROLE	6 zemí EU a Čína	Technická platforma pro e-learningové on-line kurzy se zaměřením na různé obory VŠ studia
<a href="http://www.role-project.eu/">www.role-project.eu/</a>			
Games and Learning Alliance	GALA	31 univerzit a výzkumných institucí z EU	Vývoj počítačových her s výukovým potenciálem (tzv. serious games)
<a href="http://www.galanoe.eu/">www.galanoe.eu/</a>			
Immersive Reflective Experience-based Adaptive Learning	ImREAL	7 evropských univerzit	E-learningový systém pracující se sociálními sítěmi a využívající prvky zážitkové pedagogiky
<a href="http://www.imreal-project.eu/project.html">www.imreal-project.eu/project.html</a>			

Tabulka 2.1: Zajímavé současné nástroje využívající principy počítačového adaptivního testování. Prvních 7 nástrojů převzato od Filípkové a Byčkovského (2008), doplněné dalšími.



## 2.2 Přínosy počítačového adaptivního testování

---

### 2.2 Přínosy počítačového adaptivního testování

Přínosy počítačového adaptivního testování můžeme rozdělit do dvou velkých skupin, které se však mohou v některých situacích částečně překrývat. Z pohledu zadavatele testu je výrazným přínosem celkové zkrácení testů, neboť nebudou administrovány žádné položky, jež jsou pro daného klienta zbytečné. Výzkumy opakovaně prokázaly, že počítačové adaptivní testy bývají průměrně o 50% kratší než jejich klasické verze ve formě tužka-papír při zachované, nebo dokonce zlepšené přesnosti měření (Embretson, Reise, 2000). Dalším pozitivem je také možnost opakovaného testování (například administrace testu před a po nějakém terapeutickém zásahu, nebo na začátku jazykového kurzu a po jeho absolvování apod.). Můžeme snadno zajistit, aby respondent odpovídal na jiné položky, a tudíž se vyvarovat efektu zapamatování si odpovědí.

Z pohledu testanta je výrazným přínosem především tzv. testování šité na míru jednotlivci (angl. tailored tests). Tento způsob testování totiž výrazně zlepšuje motivaci testanta ke spolupráci, a tudíž snižuje neúspěch v testu zapříčiněný jinými vlivy než schopností, kterou test měří. V rámci klasických testů obvykle klient odpovídá na mnoho otázek, které nejsou příliš přiměřené jeho schopnostem. Pokud je proband v rámci testu nucen odpovídat na otázky, které jsou pro něj příliš složité a obtížné, může u něj vzrůstat obava z dalších otázek a nechuť na ně odpovídat. Pokud naopak proband opakovaně odpovídá na pro něj příliš jednoduché až triviální otázky, obvykle se začíná nudit a opět o test ztrácí zájem (odmítne test dokončit, dělá chyby z nepozornosti, protože se soustředí na něco jiného apod.). Optimální úroveň motivace k řešení testu tak klient dosáhne, pokud jsou v souladu nároky testu a jeho schopnosti, v ideálním případě by se

## Počítačové adaptivní testování

---

mohl přiblížit prožitku flow (Csikszentmihalyi, 1996), kdy zapomíná na okolí a je zcela ponořen do zadaného úkolu. Ideální jsou tedy takové položky, u nichž je zhruba 50% pravděpodobnost, že je klient s danou úrovní měřené schopnosti odpoví správně.

Hambleton, Swaminathan a Rogers (1991) výhody počítačového adaptivního testování stručně shrnují v následujících bodech:

- zvýšení bezpečnosti testování (ve smyslu úniku informací)
- není potřeba testových archů
- individuální tempo je voleno respondentem
- okamžitá znalost testového skóru, na niž navazuje bezprostřední zpětná vazba klientovi
- minimalizace testové frustrace pro většinu respondentů
- větší soubor pro standardizaci testu
- možnost jednoduchého odstranění nefunkčních položek z položkové banky
- větší flexibilita ve výběru formátu položek

Filípková a Byčkovský (2008) k těmto výhodám ještě přidávají možnost zabránění opisování (i pokud bude test zároveň v počítačové učebně plnit více studentů, nemohou vzájemně opisovat, neboť každý odpovídá na jiné otázky), možnost využití celé řady multimediálních prvků, z nichž mohou být tvořeny interaktivní položky, a možnost okamžitého vyhodnocení a interpretace testu. Poslední dvě výhody jsou však platné i pro klasické testy administrované počítačem a nesouvisí přímo s adaptivitou testu. Zajímavou variantou, na kterou

## 2.3 Průběh počítačového adaptivního testování

---

Filípková a Byčkovský (2008) upozorňují, je také možnost generování „klasických“ testů z CAT, pokud potřebujeme připravit testy pro probandy, o jejichž schopnostech v dané oblasti máme dopředu přibližnou informaci. Např. u testu jazykových schopností, můžeme předpokládat, že žáci v určitém ročníku (respektive po určitém počtu odstudovaných let cizího jazyka) budou dosahovat konkrétní úrovně. Žáci se samozřejmě budou v rámci třídy lišit ve svých schopnostech, nicméně tento rozdíl bude nižší než u žáků napříč celou školou. Na základě CAT tak můžeme vygenerovat odpovídající testy s obtížností pro páté či osmé třídy apod.

## 2.3 Průběh počítačového adaptivního testování

Příprava a průběh počítačového adaptivního testování se skládá z několika fází, které si nyní přiblížíme.

### 2.3.1 Položková banka

Základním stavebním kamenem počítačového adaptivního testování je položková banka. Jedná se o množinu položek, z níž jsou jednotlivé položky při administraci vybírány. Poznamenejme, že test bude tak dobrý a úspěšný, jak vhodná a podařená bude položková banka. Počet položek v této bance musí být velký, minimálně v řádech stovek, v ideálním případě i tisíců (Hambleton, Swaminathan, Rogers, 1991). V rámci této množiny jsou samozřejmě známy vlastnosti každé položky, případně jejich vzájemné vztahy. Například v kontextu teorie odpovědi na položku (o níž bude podrobněji pojednáno dále) potřebujeme položky s vysokou rozlišovací schopností, které probíhají celým

## Počítačové adaptivní testování

---

spektrem obtížnosti a mají minimální parametr uhádnutelnosti.

Vytvořit optimální položkovou banku je obtížné právě kvůli potřebě vysokého počtu vhodných položek. Obzvláště v oblasti osobnostních testů, které měří konstrukty jako např. agrese, úzkost atd., se málokdy najdou škály se sto položkami. I z tohoto důvodu je počítačové adaptivní testování stále více rozšířeno spíše v rámci administrace výkonových testů (Embretson, Reise, 2000).

### 2.3.2 Začátek testování

Když máme vybudovanou kvalitní položkovou banku, následuje logicky otázka, jak začít testovat. Jakou položku vybrat jako první? Můžeme-li předpokládat normální rozložení zjišťované schopnosti, doporučuje se zadat několik prvních položek (obvykle 3 až 5) s parametrem obtížnosti zhruba  $-0,5$  až  $0,5$  (Embretson, Reise, 2000). Pokud máme o daném respondentovi nějaké dřívější reference, můžeme tyto informace využít při výběru prvních položek. Ovšem i zcela náhodně zvolené první položky vedou při použití dostatečně rozsáhlé a rozmanité položkové banky k užitečným výsledkům. Problém není v technických otázkách, je však třeba zohlednit určitý psychologický přístup ke klientům. Sami bychom se proto přikláněli spíše k snadnějším položkám, než ke zcela náhodnému výběru, neboť nesmíme zapomínat na důležité psychologické faktory, jimiž jsou lidská motivace k práci s testem a snaha po minimální testové anxietě.

Po zodpovězení dvou až tří prvních položek program obvykle vyhodnotí vstupní úroveň schopnosti a začne vlastní výběr položek pro daného respondenta.

## 2.3 Průběh počítačového adaptivního testování

---

### 2.3.3 Vlastní testování

V této fázi je nutné použít nějakou proceduru pro výběr položek. Po každém zodpovězení položky respondentem se znovu určuje úroveň respondentovy schopnosti a podmíněná standardní chyba. Na základě těchto údajů program vybere a předloží další položku.

V současnosti se pro výběr položek používají dvě procedury. První z nich je založena na principu maximálního přínosu informace (Hambleton, Swaminathan, Rogers, 1991). Vybere se taková položka, která přináší největší informaci, a tudíž minimalizuje standardní chybu při odhadnuté úrovni schopnosti respondenta. Samozřejmostí je vyvarovat se toho, aby se některá položka administrovala vícekrát.

Jinou metodu představuje Bayesovská volba položky. Tato procedura vybírá položku tak, aby byl zajištěn co nejmenší rozptyl výsledného rozložení respondentovy schopnosti. S každou další zodpovězenou položkou je rozložení užší, takže dosahujeme větší přesnosti při odhadu úrovně respondentovy schopnosti.

### 2.3.4 Ukončení testu

Další otázka související s počítačovým adaptivním testováním zní, kdy toto testování ukončit. Můžeme si stanovit určitou míru přesnosti, se kterou chceme měření provést. Potom test skončí v okamžiku, kdy bude dosaženo předem stanovené hodnoty podmíněné standardní chyby. Také se dá přestat tehdy, když odpovědi na další položky již nebudou mít podstatný vliv na klesání podmíněné standardní chyby.

Existují i počítačové adaptivní testy s předem určeným fixním počtem položek, domníváme se však, že tím přicházejí o část své adaptivity, a proto toto řešení nepreferujeme.

## 2.4 Matematické teorie stojící v pozadí CAT

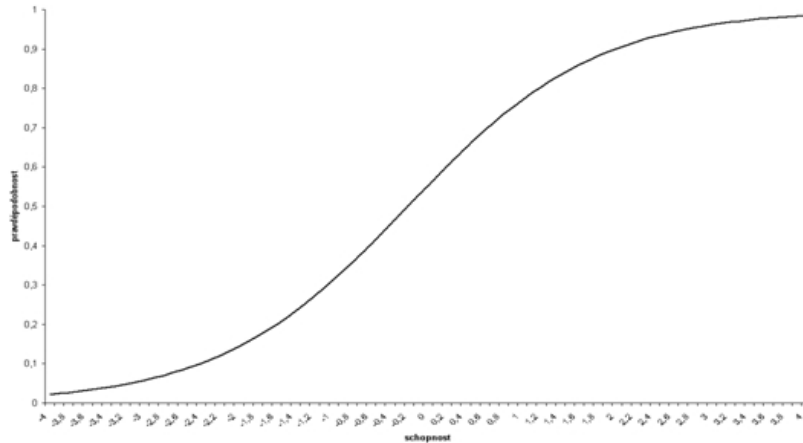
Nejrozšířenější teorie, kterou využívá CAT je v současné době bezesporu teorie odpovědi na položku (dále IRT, z anglického Item Response Theory). Než se budeme podrobněji věnovat této teorii, z které vychází i adaptivní test COMPACT, chtěli bychom zdůraznit, že se nejedná o jedinou možnou matematickou teorii (ač se o tom někteří autoři snaží veřejnost přesvědčit). Dalšími, kvantitativně méně se vyskytujícími, zato však eliminující některá kvalitativní omezení IRT (hlavně předpoklad jednodimenzionality, který je v reálu obtížně splnitelný) jsou teorie vědomostního prostoru (dále KST, z anglického Knowledge Space Theory) a teorie měření rozhodování (dále MDT, z anglického Measurement Decision Theory).

### 2.4.1 Základní principy teorie odpovědi na položku

Dva základní předpoklady IRT teorie jsou (Hambleton, Swaminathan, Rogers, 1991):

- Výkon respondenta na testové položce je predikovatelný (vysvětlitelný) množinou faktorů nazývanou rysy, latentní rysy nebo schopnosti.
- Vztah mezi výkonem respondenta na položce a množinou rysů, jež tento výkon zapříčiňují, může být popsán monotónní rostoucí funkcí nazývanou charakteristická funkce položky (item characteristic function). Tato křivka má tvar normální ogivy.

## 2.4 Matematické teorie stojící v pozadí CAT



Obrázek 2.1: Charakteristická funkce položky v rámci IRT

Společným předpokladem IRT modelů je to, že množina položek (tedy celý test nebo subtest) měří pouze jednu schopnost. Tato podmínka samozřejmě není v reálu nikdy zcela splněna, jde spíše o ideál, k němuž se při výzkumech i jiných aplikacích snažíme co nejvíce přiblížit. Do hry vždy vstupuje množství více či méně odhalitelných faktorů jako např. ochota spolupracovat, testová úzkost, úroveň motivace, schopnost pracovat rychle, tendence ke kladným odpovědím atd. Důležité tedy je, aby se daná měřená schopnost nerozplynula v těchto ostatních vlivech, jež není možné odstranit, ale aby byla hlavním faktorem, který ovlivňuje výsledný testový skóre. Dalším důležitým předpokladem je předpoklad lokální nezávislosti. Odpovědi zkoušeného na každé dvě položky jsou statisticky nezávislé, což znamená, že neexistuje žádný vztah mezi odpověďmi respondenta na různé položky. Tento předpoklad částečně nahrazuje požadavek jednodimenzionality, jehož absolutní splnění je nemožné. Předpoklad lokální nezávislosti nám pomáhá při tvorbě modelu uvažovat právě nad těmi schopnostmi, které opravdu ovlivňují odpovědi na položky. Po-

## Počítačové adaptivní testování

---

kud je splněn předpoklad jednodimenzionality, je zároveň splněna i lokální nezávislost.

IRT modely, jež vychází z výše uvedené podmínky, se nazývají jednodimenzionální. Nyní se objevují už i vícedimenzionální modely, nicméně jejich snaha vymanit se z omezení jednodimenzionality nebývá vždy korunována úspěchem.

### 2.5 IRT modely

Postupně bylo vytvořeno mnoho modelů, které popisují charakteristickou křivku položky při respektování výše uvedených předpokladů. Tyto modely se mezi sebou liší počtem parametrů, pomocí nichž je charakteristická křivka popisována.

Nejjednodušším IRT modelem je jednoparametrový logistický model. Matematické vyjádření je následující:

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} \quad \text{pro } i=1, 2, \dots, n$$

- $P_i(\theta)$  je pravděpodobnost, že náhodně vybraný respondent s úrovní schopnosti  $\theta$  odpoví na položku  $i$  správně;
- $b_i$  je parametr obtížnosti položky  $i$ ;
- $n$  je počet položek testu;
- $e$  je Eulerovo číslo.

Parametr položky  $b_i$  je bod na škále schopností, v němž je pravděpodobnost správné odpovědi rovna 0,5. Čím vyšší je hodnota  $b_i$ , tím



## 2.5 IRT modely

---

větší schopnost je požadována po respondentovi, aby pravděpodobnost jeho správné odpovědi byla 50%, a tím je tedy položka obtížnější. Jednodušší položky jsou na schopnostní škále situovány do levé nebo spodní části, zatímco obtížnější položky se vyskytují v pravé nebo horní oblasti.

Tento model je vhodný v situacích, kdy můžeme předpokládat, že položky v testu mají přibližně stejnou míru rozlišovací účinnosti.

Dvoupřímý logistický model vyvinul Birnbaum. Podobá se jednopřímému modelu uvedenému výše, obsahuje však dva další elementy.  $D$  je konstanta, která formátuje logistickou funkci, aby byla tvarově co nejvíce podobná normální ogivě; má hodnotu 1,7. Druhým přidaným elementem je parametr  $a_i$ , nazývaný diskriminační parametr. Vyjadřuje velikost naklonění charakteristické křivky položky v bodě  $b_i$ . Položky, které jsou v bodě  $b_i$  strmější (a tedy je pro ně parametr  $a$  vyšší), mají větší rozlišovací potenciál, takže jsou vhodnější pro třídění respondentů podle odlišných úrovní schopnosti  $\theta$ , než položky pozvolnější.

Diskriminační parametr je teoreticky definován v intervalu  $(-\infty, +\infty)$ . V psychometrické praxi jsou však záporné hodnoty vyloučeny. Jsou pouze indikátorem toho, že daná položka je vadná, neboť nemá smysl uvažovat o tom, že pravděpodobnost správné odpovědi roste se snižujícími se schopnostmi respondenta. Také je velmi neobvyklé získat hodnotu parametru  $a$  vyšší než 2. Diskriminační parametr se tedy pohybuje v rozmezí od 0 do 2 (Hambleton, Swaminathan, Rogers, 1991).

Dvoupřímý model je vhodný v případech, kdy odpověď na testové položky není uhádnutelná. Například pro osobnostní dotazníky (kde žádná odpověď není nesprávná), ve kterých vycházíme z předpokladu, že respondent netuší, co odpovědí na danou položku zjišťujeme, a tedy nemá možnost své odpovědi záměrně zkreslovat.

---

## Počítačové adaptivní testování

---

V tříparametrovém logistickém modelu jsou zachovány výše popsané parametry  $a_i$ ,  $b_i$ , a objevuje se třetí parametr  $c_i$  nazývaný parametr uhádnutelnosti (v originále pseudo-chance-level parameter nebo guessing parameter). Vyjadřuje pravděpodobnost toho, že respondent s nízkou schopností odpoví na položku správně. Parametr uhádnutelnosti má hodnotu pravděpodobnosti, s jakou i respondent s nulovou měřenou schopností „vyřeší“ položku správně. V praxi se objevuje ve formátech položek s vícenásobnou volbou. Pokud je testovaným osobám předložen výběr z  $n$  možností (jedna správná odpověď a  $n-1$  distraktorů), je pravděpodobnost správné odpovědi bez zapojení měřené schopnosti  $1/n$ . To navíc pouze v případě, že všechny nabízené distraktory jsou na první pohled rovnocenné (občas se mohou zvláště v oblasti pedagogického testování objevit situace, kdy třeba ze čtyř nabízených možností jsou dvě zřetelně nesmyslné, a proband pak tipuje správnou odpověď pouze ze zbývajících dvou nabídek). Je zřejmé, že při tvorbě testových položek by měla převládat snaha o minimalizaci parametru uhádnutelnosti. Tříparametrový model je vhodný k vývoji pedagogických testů. Podrobnější informace o modelech IRT lze nalézt v Urbánek, Denglerová, Širůček (2011) nebo Denglerová (2007).

## 2.6 Informační funkce položky v kontextu IRT

Tato funkce má velký význam při konstrukci testů, neboť právě velikost informace, kterou daná položka přináší, může být vhodným kritériem pro rozhodování, zda položku ponechat, nebo ji vyřadit z testu. Současně se jedná o analogii reliability z klasické teorie testů.

## 2.6 Informační funkce položky v kontextu IRT

---

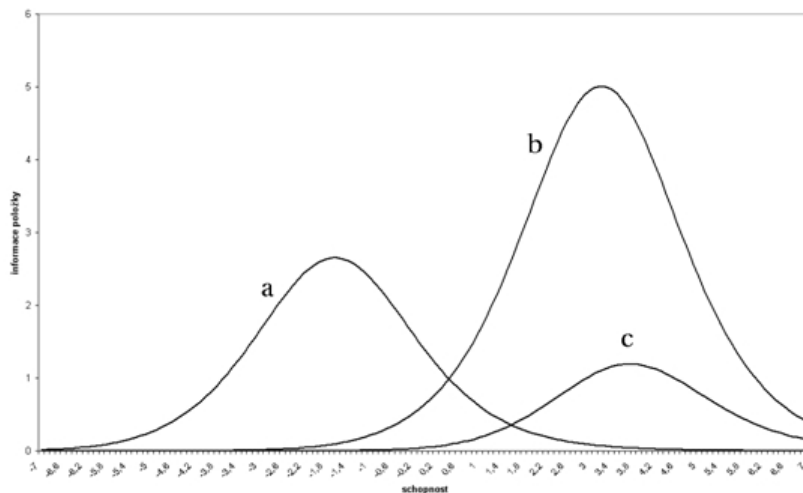
$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad \text{pro } i=1, 2, \dots, n$$

- $I_i(\theta)$  je velikost informace, kterou poskytuje položka  $i$  při úrovni schopnosti  $\theta$ .
- $P_i(\theta)$  je pravděpodobnost správné odpovědi.
- $P'_i(\theta)$  je derivace  $P_i(\theta)$ .

Z výše uvedeného vzorce vyplývá, že informační funkce položky souvisí s parametry konkrétní položky a také s úrovní schopnosti  $\theta$ . Nyní se podíváme na tyto souvislosti podrobněji:

- Větší množství informace poskytují položky s vyšší obtížností.
- Diskriminační parametr položky podstatně ovlivňuje velikost informace, kterou daná položka poskytuje. Čím je diskriminační parametr vyšší, tím větší má položka informační hodnotu. Položky s nízkým diskriminačním potenciálem jsou v rámci testu statisticky zbytečné. Mohou však mít význam například na začátku testu jako zácvičné položky, neboť je vhodné, aby je zvládla většina respondentů a neztratila tak motivaci pro další práci s testem.
- Se zvyšující se hodnotou parametru uhádnutelnosti samozřejmě informační hodnota klesá, neboť i ti respondenti, kteří nedisponují danou schopností, mají jistou pravděpodobnost (dle velikosti parametru  $c$ ), že na položku správně odpoví.
- Informační hodnota položky se různí dle úrovně schopností. Položka s relativně vysokou obtížností má tedy velkou informační

## Počítačové adaptivní testování



Obrázek 2.2: Informační funkce položek

hodnotu mezi respondenty s vysokou mírou dané schopnosti, ale ve střední oblasti schopnosti nám tolik informace neposkytne.

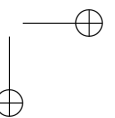
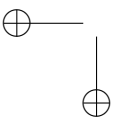
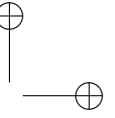
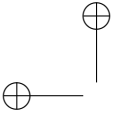
Jako charakteristiku celého testu můžeme používat informační funkci testu, což je součet informačních funkcí všech položek, které test obsahuje. Z toho plyne, že konkrétní položka přispívá testu zcela nezávisle na celkové informační funkci testu. To znamená, že příspěvek konkrétní položky můžeme zjistit, aniž bychom znali ostatní položky testu.

Výše jsme si objasnili, že užitečnost konkrétní položky ke správnému odhadu latentního rysu, potažmo celého testu, závisí také na úrovni schopnosti daného respondenta. Pokud tedy administrujeme určitý soubor položek skupině respondentů, výsledky, které obdržíme, nebudou stejně přesné pro každého z nich. Při nějaké úrovni schopnosti přináší tyto položky největší informaci, a tudíž je minimalizována standardní chyba. Pokud respondent s vysokou mírou schop-

## 2.6 Informační funkce položky v kontextu IRT

---

nosti bude testován položkami s nízkou obtížností, zřejmě je zodpoví všechny správně, a v rámci testu nezískáme představu o tom, kde je strop jeho schopnosti. Ideálním řešením by tedy bylo, aby každý respondent byl testován jedinečnou množinou položek, jež nejvíce odpovídá jeho schopnostem. Nejvhodnější jsou položky, u nichž míra pravděpodobnosti správné odpovědi variuje kolem 50%. To znamená, že pro každého respondenta jsou nejvhodnější ty položky, jejichž obtížnost odpovídá úrovni jeho měřeného atributu. Takovýto přístup k testování, tzv. testy šité na míru jednotlivci (angl. tailored tests), je základem adaptivnosti.



## Kapitola 3

# Vznik a dosavadní vývoj adaptivního testu COMPACT

Tato kapitola se věnuje důvodům, jež vedly k zařazení tvorby adaptivního testu do projektu COMPACT a vyplývají z dlouhodobé potřeby zajistit odpovídající jazykovou úroveň absolventů MU. Dále pak popisuje vývoj tohoto nástroje a shrnuje zkušenosti z několika semestrů nasazení. Poslední část kapitoly předkládá technické parametry prostředí.

### 3.1 Východiska

Výukou jazyků, především pak angličtiny pro akademické účely, musí na MU, stejně jako na všech ostatních vysokých školách České republiky, každoročně projít tisíce studentů. Často je jich víc, než dovoluje kapacita jednotlivých seminárních skupin, a proto je nutno vybrat ty,

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

kteří mají nejlepší předpoklady, aby předmět skutečně i úspěšně dokončili. Celouniverzitním požadavkem je, aby studenti bakalářského, magisterského i doktorského studia povinně absolvovali zkoušku z cizího jazyka se stanovenou minimální výstupní úrovní podle Společného evropského referenčního rámce pro jazyky (SERRJ), a to konkrétně B1 pro bakalářské studium, B2 pro magisterské studium a C1 pro doktorské studium, přičemž výuka jazyka je koncipována v každém cyklu většinou jako dvousemestrová s hodinovou až čtyřhodinovou dotací týdně podle typu fakulty.

Z tohoto hlediska je jen logické, že fakultní pracoviště CJV MU trvají na minimální vstupní úrovni, které musejí studenti dosáhnout, aby si mohli semináře odborného cizího jazyka zapsat. Za dva maximálně patnáctitýdenní semestry nijak zvlášť intenzivní výuky, která je navíc na žebříčku studijních povinností ve srovnání s vlastním oborem přece jen sekundární, se od studentů žádné zázraky očekávat nedají. V průměru je podle odhadů Cambridge ESOL k dosažení jednotlivých úrovní zapotřebí určitého, a to nijak zanedbatelného počtu hodin řízené výuky (guided learning hours). K dosažení A2 je zapotřebí přibližně 180-200 hodin, k dosažení úrovně B1 asi 350-400 hodin, B2 500-600 hodin, C1 700-800 hodin atd. (ALTE Can Do statements and the CEFR). Má-li tedy student pokročit od B1 v bakalářském studiu na B2 v magisterském, potřebuje k tomu odhadem 150 až 200 hodin za předpokladu, že úrovně B1 skutečně dosáhl a má za sebou alespoň 350 hodin výuky angličtiny v předchozím studiu. Pokud se student na začátku dvousemestrového studia odborné angličtiny pohybuje někde mezi A1 a A2, pak většinou zkoušku z odborného jazyka na úrovni B1 v bakalářském a na úrovni B2 v magisterském studiu není schopen zvládnout.

Ani plošné zavedení povinné státní maturity žádný výrazný pokrok v tomto směru nepřineslo. Poptávka jednotlivých pracovišť CJV MU



### 3.1 Východiska

---

se totiž nemůže setkat s plošnou nabídkou absolventů středních škol: angličtina není jediný cizí jazyk, z něhož lze maturovat, některé fakulty však jiný cizí jazyk nenabízejí. Navíc nelze automaticky předpokládat, že každý, kdo odmaturoval z angličtiny skutečně dosahuje úrovně deklarované pro nové státní maturity. Žák by podle informací MŠMT měl na základní úrovni maturitní zkoušky z cizího jazyka dosáhnout B1, na vyšší pak B2. Výsledky podzimního běhu našeho testu tuto úroveň jako minimální nepotvrzují. Odhlédneme-li od 126 čerstvých maturantů, kteří dnes studují na LF MU a úrovně B1 nedosahují (může jít o maturanty z jiných jazyků, ale na LF se vyučuje pouze angličtina), situace u 124 studentů FF a FSS je jiná: možnost zvolit si jiný jazyk měli, a přesto si zvolili angličtinu, takže z ní pravděpodobně i maturovali, přitom však úrovně B1 nedosahují.

Tyto výsledky nejsou samozřejmě dostatečně průkazné, aby se z nich daly vyvozovat relevantní závěry. Teprve dlouhodobějším sledováním jazykové úrovně čerstvých maturantů v prvních ročnících MU bude možné potvrdit, zda studenti středních škol dosahují při absolvování deklarované úrovně státních maturit, ale jisté je, že skupina dvou set čerstvých maturantů, které jsme měli příležitost naším nástrojem otestovat během prvního měsíce jejich studia na jedné z brněnských vyšších odborných zdravotnických škol dopadla podstatně hůř než studenti MU.

Sami studenti si vesměs nejsou těchto souvislostí vědomi a většinou ani netuší, v čem spočívají jejich nedostatky v cizím jazyce. Musí být proto včas informováni, jaká je jejich úroveň a v čem konkrétně mají mezery proto, aby si své znalosti – pokud už je nezískali na střední škole – doplnili.

Ať z toho či onoho důvodu, vstupnímu testování jazykových znalostí se zřejmě ještě dlouho nevyhneme. I v ideálním případě, že úroveň maturantů bude skutečně odpovídat B1 a B2 podle SERRJ a cel-

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

kový počet vysokoškolských studentů v rámci demografického vývoje klesne natolik, že do seminářů se dostanou všichni, pak asi zase budeme jako učitelé chtít, aby seminární skupiny byly homogenní a nám i studentům se v nich dobře učilo. V tom případě už stojí za to do vstupního testování centrálně investovat čas a energii.

Vstupní testy se na jazykových pracovištích MU používaly odjakživa, aby se eliminovali studenti, kteří si na vysokou školu nepřinesli dostatečné znalosti ze středních škol, nebo od jejich maturity na střední škole již uplynula spousta let a oni nevěnovali odpovídající pozornost udržení a rozvoji jazykových dovedností. Studentům tyto vstupní testy většinou nedávaly dostatečnou zpětnou vazbu, na co by se při svém dalším studiu měli zaměřit. Pro učitele představovala jejich tvorba, administrace a vyhodnocování časovou i organizační zátěž. Navíc tyto testy byly často nestandardizované a v žádném případě nebyly, ani nemohly být, adaptivní. Každý student dostal jednu z vypracovaných variant testu, jehož úroveň byla často stanovena pouhým odhadem, byť expertním. Nedalo se stoprocentně zajistit, aby obtížnost jednotlivých variant byla identická, ale zároveň nemohla existovat pouze jedna jediná varianta, která by se používala opakovaně pro všechny. Riziko postupného prozrazení testových položek by bylo totiž příliš vysoké. Navíc tyto varianty při nejlepší vůli autorů nemohly být koncipovány jinak než lineárně podle klasické testové teorie, což mělo mimo jiné za následek, že nejpřesnějších výsledků dosahovali ti studenti, jejichž úroveň testovaných znalostí byla v pásmu průměrných výkonů dané referenční skupiny. Přidáme-li k tomu jistou míru pravděpodobnosti, že položky neprošly odpovídajícím procesem tvorby standardizovaných testů, pak několik nedostatečně validních úloh mohlo celý výsledek výkonových testů výraznou měrou negativně ovlivnit (Filípková, Byčkovský, 2008). Můžeme samozřejmě namítnout, že se jednalo pouze o vstupní testy, na kterých zase až tolik nezáleží, pokud student nakonec při výstupu u zkoušky

### 3.1 Východiska

---

své znalosti prokáže. Ale jak zajistit, aby se při výstupu skutečně jednalo o odpovídající úroveň SERRJ, jehož rámcové deskriptory jednotlivých úrovní jsou zatím poměrně vágní?

Dokud bylo nutné jen vybrat ze zaregistrovaných studentů odpovídající počet podle kapacity seminárních skupin, bylo možné stanovit hranici mezi úspěšností a neúspěšností na základě tohoto čísla. Problém se ale prohloubil přechodem na vstupní a výstupní požadavky vyjádřené škálou SERRJ, na jejichž změření však chyběl nástroj. Pokud si pracoviště vytkne za cíl, že jeho absolventi budou na určité úrovni podle SERRJ, je nutné takový nástroj, který tuto úroveň potvrdí, zajistit. Jak jsme již vysvětlili výše, při omezeném počtu hodin výuky angličtiny nelze předpokládat, že student vykazující na začátku studia např. úroveň A1 nebo i nižší se může během dvou semestrů studia posunout na úroveň, která je jasně daná pro celou univerzitu. Při hledání řešení se nabízely jen dvě cesty – vyzkoušet a zvážit již existující diagnostické testování online, například Dialang, nebo se pustit do tvorby vlastního testovacího nástroje. Finanční náročnost komerčních testů a mezinárodních certifikátů vylučovala třetí možnou cestu – objektivně si nechat studenty otestovat třetí nezúčastněnou stranou.

Ambiciózní mezinárodní projekt Dialang však našim potřebám nevyhovoval, a to hned z několika důvodů. Tím prvním je neobratná a neintuitivní manipulace. Kromě toho, že ani pod záštitou Lancaster University se od roku 2003 v podstatě nerozvíjí z důvodu nedostatku financí a samotná univerzita přiznává, že její technici jsou schopni ve svém volném čase nanejvýše zajistit jeho holou existenci, pro absolvování testu je nutné si na počítač nepříliš uživatelsky přátelským způsobem instalovat klienta. To by studentům komplikovalo život a neumožňovalo by jim to přístup z počítačových učeben MU. Další nevýhodou je skutečnost, že pokyny jsou uvedeny pouze ve čtr-

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

nácti jazycích států, které se do tohoto projektu zapojily. Instrukce v angličtině by začátečníkům dělaly potíže. Výsledky testu dále lze tisknout pouze z obrazovky, a tudíž by jejich dokládání při registraci a zápisu do seminární skupiny opět představovalo administrativní zátěž. S adaptivitou nemá tento nástroj společného nic. Pouze testantům předkládá jednu ze tří variant statického lineárního testu na základě jejich sebehodnocení a poměrně hrubého rozřazovacího testu, které spočívá v tom, že vybírají ze seznamu slov ta, která poznávají, a jsou penalizováni za slova, jež „poznali“, ale přitom v angličtině neexistují. Tento postup, který nám připadal zajímavý, jsme se také pokoušeli zavést v počáteční fázi, kdy ještě existovaly gramaticko-lexikální kategorie a položky se generovaly nikoli podle své informační hodnoty, ale v rámci těchto kategorií. Snažili jsme se tak vyřešit jeden z problémů adaptivního testování: jak nastavit vstupní úroveň, jakou obtížnost by měla mít úplně první otázka a pár následujících tak, aby testant zbytečně dlouho nebloudil položkami neadekvátními ve vztahu ke své skutečné úrovni znalostí. Nakonec jsme našli lepší řešení, o kterém bude řeč v další části této kapitoly.

Po neúspěšném hledání vhodného nástroje bylo nakonec do projektu COMPACT zařazeno vytvoření vlastního testovacího prostředí pro rychlé, komplexní a objektivní změření vstupní úrovně prostřednictvím adaptivního testu, který by na jedné straně rozřazoval studenty podle úrovně SERRJ, na straně druhé by studentům napověděl, na co by se měli ve svém dalším studiu zaměřit. Učitelům by pak nabízel otevřený, snadno rozšiřitelný nástroj pro tvorbu a distribuci testů se zabudovaným statistickým aparátem. V rámci tohoto nástroje by mohli z banky položek například vygenerovat studentům klasický test nejen s položkami na ověření zvládnutí specializované slovní zásoby, ale měli by zároveň jistotu, že student skutečně dosahuje odpovídající úrovně podle SERRJ.

## 3.2 Vývoj testu

---

### 3.2 Vývoj testu

Předem je nutno konstatovat, že naším cílem nebylo vytvořit spolehlivý test pro objektivní zařazení studentů do všech šesti zmiňovaných úrovní SERRJ. Již bylo řečeno, že žádný test není schopen pouze na základě gramaticko-lexikálních položek ve formátu výběru z několika odpovědí nebo krátkých otevřených odpovědí přesně oddělit úroveň C od úrovní B, není schopen určit úroveň samostatného používání jazyka. Vychází to z popisu úrovní v oficiálním dokumentu SERRJ.

Uživatel úrovně C1 a výše musí prokázat znalost gramatiky jako součást ostatních dovedností (Reichová et al, 2010), musí být schopen se plynule a pohotově vyjadřovat bez zjevného hledání výrazů, musí umět užívat jazyk pružně a efektivně pro nejrůznější společenské, akademické a profesní účely. Musí být schopen vytvořit srozumitelné, dobře uspořádané texty na složitá témata.

Charakteristika uživatele úrovně C2 říká, že ten, kdo jí dosahuje, by měl porozumět téměř všemu, co si vyslechne nebo přečte, měl by umět shrnout informace z různých mluvených a psaných zdrojů a zároveň by měl umět tyto informace přednést v logicky uspořádané podobě (Common European Framework of Reference for Languages: Learning, Teaching, Assessment). Nejedná se tedy pouze o zvládnutí běžných jazykových prostředků, ale i o kompetence, jichž s velkou pravděpodobností spousta lidí z řad rodilých mluvčích není schopna.

S ohledem na zadání, které popisujeme v první části této kapitoly, bylo naším hlavním cílem oddělit od sebe navzájem úrovně A a B a zajistit, aby v těchto intervalech test měřil přesně. Zobrazovat výsledky předpokládané úrovně C2 ani nemělo smysl, proto jako nejvyšší možnou úroveň uvádíme C1 plus. Výsledky C1 a C1 plus pouze naznačují, že převážně pasivní znalosti gramaticko-lexikálních jednotek směřují k úrovni C, která je na některých fakultách nastavena

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

jako cílová i pro magisterské studenty a počítá se s ní u všech doktorských studentů.

Vytvořit kvalitní jazykový test je časově velmi náročná činnost. Nezáleží na tom, zda jde o měřicí nástroj administrovaný ve formátu papír-tužka s následnou individuální opravou učitelem, nebo jde o částečně zautomatizovaný proces, kdy studenti řeší položky do formulářů, jež se pak opravují centrálně za pomoci skenovacího zařízení. Vždy se jedná o náročný úkol bez ohledu na to, zda má výsledný test odpovídat klasické testové teorii, či teorii odpovědi na položku. Pokud jednotlivé položky v rámci IRT, respektive celé testy v rámci klasické testové teorie nejsou dostatečně pretestované a odborně standardizované, nelze se s jistotou spolehnout na to, že všechny budou mít stejnou výpovědní hodnotu. I pro generování položek do lineárního testu je zapotřebí vytvořit dostatečně početnou banku citlivých a informačně přínosných testových položek a to není něco, co mohou jednotliví vyučující vytvářet na koleně každý sám pro sebe vedle svých běžných výukových povinností. Mělo by se vždy jednat o kolektivní práci, k níž je zapotřebí nejen tým jazykových a testovacích specialistů, ale i dostatečný počet testantů, tedy studentů, kteří svými průchody testy poskytují data potřebná pro rozvoj kvalitní banky položek. Univerzitní prostředí je z tohoto hlediska prostředím ideálním.

Naším cílem ale nebylo jen vytvořit modelový jazykový test, ale přímo test adaptivní, který by se dal používat opakovaně napříč všemi fakultami. Jak již bylo vysvětleno v předchozí kapitole, princip adaptivity v testování spočívá v tom, že každá další otázka se generuje podle toho, jakých testant dosáhl výsledků u otázky předchozí. Pokud student otázku zodpověděl správně, dostane otázku těžší, pokud ji zodpověděl chybně, dostane otázku lehčí. Na základě průchodu testem pak lze vypočítat jeho přibližnou úroveň.

### 3.2 Vývoj testu

---

Jako řada jiných adaptivních testů je i ten náš založený na kalibraci položek podle tříparametrového modelu IRT, odhadu schopností, který se v průběhu testu u každého studenta dále zpřesňuje, a na optimálním logaritmu generování jednotlivých položek. Jeho cílem je, aby se každému z testovaných studentů generovaly pouze takové položky, jež mají ve vztahu k jeho odhadované jazykové úrovni nejvyšší informační hodnotu – jinými slovy, je u nich padesátiprocentní pravděpodobnost správné odpovědi.

Práci na adaptivním testu jsme zahájili ještě v období přípravné fáze, abychom měli jistotu, že nebudeme pracně vymýšlet již vymyšlené. Nenašli jsme však nic, co by i jen vzdáleně připomínalo naši vizi. Samotný projekt COMPACT zahájil svou činnost v květnu 2009, nicméně díky zkušenostem s výukou online, především pak z předchozích běhů předmětu Online\_A a předmětu Procvičování anglické gramatiky online, jsme měli v té době už docela slušnou banku položek, u nichž jsme byli schopni alespoň orientačně určit obtížnost pomocí nástrojů Informačního systému MU (IS). Měli jsme tedy zárodek položkové banky, prvních asi osmnáct set položek pro případné spuštění testovacího provozu jednoparametrického adaptivního testu. Chybělo nám ovšem vhodné prostředí, protože e-learningová agenda ve formátu odpovědníků v IS nic takového nenabízí.

První dvě pilotní testovací vlny probíhaly stále ještě v IS, kde bylo možné adaptivitu nasimulovat pouze tak, že se náhodně generovaly položky z různých, předem navržených gramaticko-lexikálních a obtížnostních kategorií, což nelze považovat za plnohodnotné pretestování pro účely adaptivního testu. O obtížnosti, která pro nás byla v dané fázi projektu vlastně jediným parametrem, jsme u řady položek měli jakési povědomí také proto, že pocházely z předchozích verzí vstupních testů a ze studijních materiálů výše zmiňovaných předmětů vyučovaných online. Zatímco jsme sbírali data o položkách

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

v IS, bylo nutno udělat podrobnou analýzu potřeb pro vytvoření optimální verze vlastního softwaru tak, aby byl schopen provádět potřebné matematické operace nutné pro kalibraci, nabízel přehlednou a intuitivní manipulaci s položkami i s celými testy, včetně grafického vyjádření charakteristické funkce položky v rámci IRT, byl multimediální a umožňoval snadnou správu uživatelů a převod jejich výsledků do IS. Mezitím bylo nutno sbírat, editovat a kategorizovat nové testové položky a ukládat je do položkové banky.

Jakmile bylo prostředí na serveru RMU naprogramované a vyladěné, vytvořili jsme v něm čtyřicet gramatických kategorií a naplnili je hrubě pretestovanými položkami, které jsme rozdělili do šesti stupňů obtížnosti. Položky byly kalibrovány jen na základě vypočítaného indexu obtížnosti z pilotovací fáze. Pak byl nutný další pretest v novém prostředí, tentokrát již v režimu, který se blížil adaptivnímu, byť do plnohodnotného měl ještě daleko. Do prostředí na rektorátním serveru byla na základě klasické položkové analýzy vybrána a importována necelá třetina z původního počtu, tedy asi dva a půl tisíce položek.

Po výpočtu parametrů vzniklo jakési kontinuum obtížnosti, citlivosti a parametru pseudohádání, došlo ke zrušení kategorií a bylo nastaveno generování testových otázek z položkové banky tak, aby po správné odpovědi následovala obtížnější, kdežto po chybné student dostal položku lehčí. K úpravě a doplňování dochází průběžně a mělo by k němu docházet i po skončení projektu.

Pro první běh v novém prostředí v podzimním semestru 2010 se test skládal ze dvou částí. První část, pro niž jsme se nechali inspirovat již zmíněným mezinárodním projektem Dialang, byla založená na rozpoznání slovní zásoby. Seznam o počtu 460 slov obsažených v tzv. Academic Wordlist jsme zdvojnásobili vytvořením neexistujících dublet a z takto vzniklé sady se studentům náhodně vygenerovalo 100 vý-



### 3.2 Vývoj testu

---

razů, u nichž měli určit, zda je znají, nebo neznají. Výsledek pak měl sloužit k rychlejšímu nastavení prvních položek.

Tato část testu se však neosvědčila. Byla zdlouhavá, nevykázala žádnou významnou korelaci, a proto byla z dalšího rozvíjení nástroje vyřazena. Jedním z důvodů byla bezesporu nešťastná volba výrazů ze seznamu Academic Wordlist, který (jak se dalo předpokládat) obsahoval mezinárodní slova, jež ani českým začátečníkům nedělala potíže. Druhá část testu pak předkládala studentům položky z odpovídající obtížnostní úrovně tak dlouho, dokud student nenarazil na hranice svých možností, nebo možností kategorie. V jednotlivých kategoriích byly položkám podle obtížnosti přiřazeny body jedna až čtyři. Následně se posunul do další kategorie. Při skutečně bezproblémovém průchodu studentům stačilo k absolvování celého testu 40 položek, ale většinou potřebovali k absolvování testu položek mnohem více. Protože se ale jednalo o výběr z několika odpovědí zaklikáváním, byl přesto test časově velice úsporný.

Pro jarní semestr 2011 jsme sice výběr slov pro úvodní část testu upravili, ale přesto vysoká míra náhody a zdlouhavé procházení desítkami výrazů studenty k smrti nudila a demotivovala. Proto jsme při dalším pilotování tuto fázi nahradili jednou velmi jednoduchou položkou „pro zahřátí“ a sérií pretestovacích položek. Výsledky pretestu se sice do celkového hodnocení jednotlivce nezapočítávají, nicméně urychlují výchozí úroveň při generování dalších, tentokrát do celkového výsledku již započítávaných položek. Tímto způsobem průběžně získáváme další kalibrované položky a test se může dynamicky rozvíjet nejen po stránce počtu, ale i druhu položek.

Do jarního semestru 2011 byly jako novinka zařazeny k pretestu tvořené odpovědi, do podzimního semestru téhož roku pak položky na poslech. Pro další rozvoj stojí za zvážení například naplnění banky položkami na porozumění čtenému již podle oborové diferenciaci,

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

nebo přiřazování, které je rovněž prostřednictvím formátu výběr z několika odpovědí snadno realizovatelné již dnes.

V poslední fázi projektu dochází k vytvoření dvou variant adaptivního testu. Jednak jde o tzv. rychlý rozřazovací test, který testantům vygeneruje jen 30 řádných plus tři pretestovací položky, jednak o delší diagnostický test, k jehož vytvoření bylo zapotřebí osadit položky orientačními štítky (tagy) s pojmenováním gramatických struktur a lexikálních okruhů, které jednotlivé položky testují. Krátký rozřazovací test neobsahuje dotaz na stupeň jistoty, s jakou testant otázku zodpověděl, protože tento aspekt nehraje při zobrazování výsledku žádnou roli.

Výstupem krátkého testu je jen bodové ohodnocení, jeho vyjádření v úrovních SERRJ a titul vhodné učebnice pro danou pokročilost. Diagnostický test obsahuje 60 položek plus deset procent pretestovacích, a protože jeho výstupem je seznam chybně vyřešených okruhů srovnávaný se stupněm jistoty deklarovaný testantem, je mezi jednotlivými položkami mezistupeň takto formulovaného dotazu.

Přestože vhodných učebnic je celá řada, vycházíme z vlastní zkušenosti a doporučujeme studentům klasickou učebnici pro samostudium Angličtina (nejen) pro samouky a publikace z řady English Grammar/Vocabulary in Use. Se správným řešením jednotlivých položek se student nesetká nikdy (i když samozřejmě při dnešních technických možnostech a studentské vynalézavosti nelze vyloučit průběžné vytváření a ukládání takzvaných screenshotů a pozdější dohledávání správných řešení po učebnicích nebo na internetu. To ale z hlediska jazykového rozvoje daného studenta může být považováno za pozitivum), protože z důvodů ochrany dat se studenti výsledky dozvídají pouze graficky a rámcově.

Ilustrační ukázky z testu jsou součástí praktické kapitoly.

### 3.3 Využívání testu

---

## 3.3 Využívání testu

Masově lze testovat jen dvakrát ročně, a to těsně před a na začátku každého semestru, přičemž jarní testování přirozeně představuje jen zlomek testantů ve srovnání s podzimem. Na většině fakult totiž začíná dvousemestrová výuka právě s podzimem a mezi semestry není rozřazování studentů do seminárních skupin aktuální. Podle informací z databáze testem do konce roku prošlo 10607 testantů, kterým bylo vygenerováno 11975 instancí testů. Přes 1300 studentů tedy test absolvovalo vícekrát než jen jednou.

Toto číslo odpovídá počtu studentů z jiných předmětů, než jsou rozřazovací testy pro zápis studentů do seminárních skupin. Snažíme se totiž jednotlivé drobné změny před velkým testováním vždy ještě dvakrát vyzkoušet na relativně menších skupinkách studentů (asi 600 každý semestr), kteří mají tento test na začátku a na konci předmětu jako povinnou součást požadavků k zápočtu. Jde většinou buď o budoucí studenty CJV MU, kteří se v online předmětech snaží dostat svou angličtinu na požadovanou úroveň, nebo o současné studenty, kteří se v angličtině chtějí zdokonalit. Případně může jít o absolventy, kteří si snaží svou úroveň alespoň udržet nejen s pomocí gramatických drillů, ale i komunikace prostřednictvím diskusních fór.

Tento pilotovací mechanismus se osvědčil jednak tím, že přece jen s těmito studenty máme užší vztah a je možné se od nich dozvědět postřehy snáze než od anonymních davů, které testem jen procházejí a ani je nenapadne, že by něco by mohlo fungovat jinak, lépe. Právě od nich máme také zpětnou vazbu rozdílu mezi subjektivním sebehodnocením a objektivním výsledkem testu. Dále je zajímavé srovnávat posun na vstupní a výstupní úrovni v rámci jednoho semestru, případně mezi semestry. Pokud se u studenta některého z těchto předmětů projeví výrazný výkyv směrem nahoru nebo dolů, je jednodušší

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

s ním vstoupit do kontaktu a zjistit příčiny – zda spočívají v nedokonalosti testu, či v lajdáckém vyplňování.

Srovnatelnost výsledků několika průchodů u jednoho testanta svědčí o spolehlivosti testu. Výjimky toto pravidlo jen potvrzovaly. Při bližším zkoumání se ale zatím vždy identifikoval problém na straně studenta.

Jistá autonomie jednotlivých pracovišť CJV MU neumožnila jednotný způsob administrace adaptivního testu pro úplně všechny studenty MU, kterých se v daném semestru rozřazování do seminárních skupin týkalo. I přes to, že některá pracoviště CJV MU se do tohoto projektu zatím nezapojila s odvoláním na skutečnost, že mají vlastní testy a vlastní způsob rozřazování studentů do seminárních skupin, přesto počet studentů z fakult, jejichž pracoviště CJV MU adaptivní rozřazovací test zadala jako povinnou prerekvizitu pro zápis studentů do jazykových seminářů, stačil na dostatečné pretestování základní sady, která v současné době představuje 1500 plně kalibrovaných položek a stejný počet položek v pretestovacím režimu. Stanovením povinné prerekvizity byla do jisté míry zajištěna motivace studentů i pravděpodobnost, že budou pracovat na optimální úrovni.

Útok na databázi jsme zatím nezaznamenali, ani jsme nezachytili pokus o únik informací. Nelze samozřejmě vyloučit, že studenti v krizových situacích budou mít tendenci podvádět. Při samostatném zpracovávání testu u počítače si nechají někým jazykově zkušenějším radit, nebo si test dokonce nechají vypracovat od někoho jiného, případně se objeví pokusy řešení položek zveřejňovat na internetu. Může jít o situace, kdy jim například hrozí neúspěšné ukončení studia proto, že nesplnili limit, nedostali se do semináře odborné angličtiny, a proto jim chybí povinná zkouška z cizího jazyka, budou mít tendenci podvádět a při samostatném zpracovávání testu u počítače si nechají někým jazykově zkušenějším radit, nebo si test dokonce

### 3.3 Využívání testu

---

nechají vypracovat od někoho jiného, případně se objeví pokusy řešení položek zveřejňovat an internetu. Tomuto úplně zabránit nelze, pokud studenti mají testy vypracovávat v soukromí.

Vynalézavost studentů je nekonečná a apelovat na jejich odpovědný přístup ke studiu angličtiny, kterou dozajista budou v dnešním globalizovaném světě pro zdárný rozvoj svého profesního života potřebovat, není vždy účinné. Vysoký počet otázek v položkové bance, skutečnost, že položky se liší test od testu i v rámci jeho opakování timentž studentem po minimálně třech měsících, a také skutečnost, že s každým průchodem dochází k postupné kalibraci dalších a dalších položek, je do jisté míry před prozrazením chrání.

Pokud by se, jak naznačí další kapitola, mělo tohoto nástroje používat při zkouškách a jejich klasifikaci, je samozřejmě pravděpodobné, že budou probíhat centralizovaně pod dohledem vyučujících v počítačových učebnách, a nikoli individuálně z kteréhokoli počítače připojeného k internetu.

Ne že by výsledky předchozích běhů adaptivního rozřazovacího testu nebyly zajímavé a poměrně stabilní, co se týče srovnání vstupní jazykové úrovně studentů jednotlivých fakult. Pro danou chvíli však nemá smysl srovnávat výsledky z dob, kdy se teprve položky sbíraly, editovaly a kalibrovaly a hledal se algoritmus jejich adaptivního generování. Teprve podzim 2011 lze brát jako začátek referenčního období a je zapotřebí více dat a hlubší analýzy pro potvrzení nebo vyvrácení prvního dojmu, kterým je, že na PrF a FSS MU jsou přijímáni jazykově nejlépe vybavení studenti.

Pro zajímavost však v následujícím oddílu uvádíme tabulku s výsledky rozřazovacího testu v porovnání mezi fakultami z podzimního semestru 2011.

## Vznik a dosavadní vývoj adaptivního testu COMPACT

### 3.4 Podzimní semestr 2011

Na základě předchozích zkušeností s relativně pozdním začátkem zápisu studentů FF a FSS (FF: CJVA.t), který každý semestr způsoboval neřešitelnou situaci, byl pro podzimní semestr 2011 vytvořen pod PdF nový předmět ADAPT\_AJ, v němž studenti mohli v klidu splnit prerekvizitu rozřazovacího testu již 6 týdnů před zahájením semestru. V předchozích semestrech byla situace skutečně kritická. Studenti sice měli rozřazovací test hotový, ale nebylo kam jeho výsledky zapsat, protože zápis do předmětů na FF začínal dva dny před zahájením semestru, tedy ve stejné době, kdy se už potřebovali zapisovat do předmětů specializované angličtiny. Teprve po zápisu, vždy po půlnoci v sobotu, se jim mohly výsledky zapsat a prerekvizitu označit jako splněnou. Data získaná z předmětu ADAPT\_AJ tedy zahrnují především studenty FF ( 1219) a FSS (424), pár zbloudilců z PdF (14) a po jednom z FI a z FSpS.

Fakulta: předmět	Test absolvovalo/zapsaných	Průměrný počet bodů	SERRJ	Počet studentů pod B1/%
ESF: BPJ_JI1	570/649	56.23	B1	55/10%
FSpS: bp939	19/83	49.47	B1	4/21%
LF: CJVAT0	726/827	56.56	B1	126/17%
PdF: ADAPT_AJ	1534/1643	57.83	B1	256/17%
z toho FF	1121	58.27	B1	117/8%
z toho FSS	411	63.15	B2	7/2%
PrF: EL003	410/481	62.5	B1/B2	15/4%
PřF: JTEST	262 /315	54.69	B1	41/17%

### 3.4 Podzimní semestr 2011

---

#### 3.4.1 Je požadavek vstupní úrovně B1 oprávněný?

Při pohledu na výsledky získané v podzimním semestru 2011 se naskýtá řada otázek a problémových okruhů, které nejsou cílem předkládané publikace, ale spolu s dalšími výsledky z následujících let mohou tvořit východiska pro budoucí výzkum. Pro shora položenou otázku a také pro posouzení validity testu je důležitým kritériem sledovat výsledky studentů v dalším období. Zajímavé proto bylo srovnat jejich výsledky z testu s výsledky hodnocení jejich semestrální práce. Bez nároků na úplnost a s vědomím, že pro výzkumné účely by bylo zapotřebí více dat, předkládáme několik kusých informací:

Ve sledovaném předmětu 73 studenti test z nejrůznějších důvodů neabsolvovali. Z těch, co ho absolvovali, 514 studentů dosáhlo minimálně úrovně B1. Z nich pak předmět do konce roku 2011 úspěšně ukončilo 281 (55%), kdežto 233 studenti do konce roku 2011 zápočet nezískali (hodnocení nezadané, neúspěšné nebo '-') (45%).

I s vědomím, že do konce roku nezískaný zápočet ještě nemusí znamenat, že studentovi se skutečně nepodařilo předmět úspěšně v daném semestru absolvovat, je zajímavé s výše uvedenou skupinou popsat skupinu těch, kdo úrovně B1 nedosáhli. Jednalo se o 56 studentů, z nichž předmět do konce roku úspěšně ukončilo jen 5 (9%), přičemž jednomu byl uznán, takže se do skupiny úspěšných nedá s čistým svědomím zařadit, zbývají tedy 4 studenti (7%). Do konce roku 2011 zápočet nezískalo 51, respektive 52 studentů (hodnocení nezadané, neúspěšné nebo '-') (91% – 93%).

Přestože jsme si vědomi, že k definitivním závěrům by bylo zapotřebí srovnat více předmětů, více semestrů a mnohem podrobněji, což by přesáhlo rozsah a poslání této publikace, na první pohled je z těchto čísel zřejmé, že pro studenty s nižší vstupní úrovní než je B1, představuje odborná angličtina vážný problém.

---

## Vznik a dosavadní vývoj adaptivního testu COMPACT

---

### 3.5 Technické parametry

Adaptivní rozřazovací test z angličtiny je webová aplikace se štihlou serverovou částí, která využívá skriptovací jazyk PHP a relační databázový systém MySQL. Serverová část v PHP je psána čistě objektovým paradigmatem s využitím jmenných prostorů. Členění aplikace je inspirováno praktikami vývoje J2EE aplikací a respektuje návrhový vzor model - view - controller. Business logika je tedy striktně oddělena od objektů sloužících pro přístup k datům. Aplikace byla vytvořena v souladu se současnými trendy v oblasti komerčního vývoje webových aplikací. Klientská část hojně využívá jazyka JavaScript, s jehož pomocí se podařilo vytvořit uživatelsky přívětivé rozhraní, které má povahu spíše desktopové aplikace než webové stránky (RIA).

Model báze dat byl vytvořen s důrazem na flexibilitu, bezztrátovost a se zohledněním vyšší zátěže během hromadného testování. Základními entitami jsou otázky, odpovědi, tagy, testy a uživatelé. Test je definován množinou otázek a jednotlivé otázky jsou charakterizovány pomocí štítků (tagů). Během testu jsou odpovědi uživatelů logovány, v době nižší zátěže systému jsou z nových odpovědí získávána data pro účely statistiky. Data jsou v databázi reprezentována formou materializovaných pohledů. Ty jsou simulovány nad běžnými tabulkami pomocí procedur, neboť materializované pohledy nejsou v MySQL implementovány. Z dat se získává především počet, kolikrát se otázka v testech objevila, úspěšnost v zodpovězení, funkčnost distraktorů, diskriminační koeficient, parametr hádání apod. Data jsou vždy dávana do kontextu s testem, ze kterého jsou získána. Lze tak pozorovat chování otázky nejen globálně, ale i v konkrétním testu a také sledovat vývoj ukazatelů v čase.

Klientská část aplikace tvoří více než dvě třetiny celkového kódu. Je využito jedinečných vlastností jazyka JavaScript, jako například



### 3.5 Technické parametry

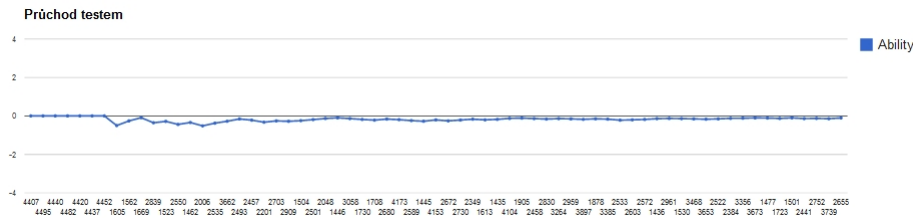
---

uzávěr a anonymních funkcí, které umožňují asynchronní styl programování, a parazitické dědičnosti Douglase Crockforda. Vzhledem k velkému množství kódu je aplikace štedře strukturovaná pomocí jmenných prostorů. Pro překlenutí rozdílů mezi webovými prohlížeči a zrychlení vývoje bylo využito JavaScriptové knihovny Dojo. Naproti tomu pro vizualizaci dat byly využity nástroje Google Charts.

Komunikace mezi klientem a serverem je realizována XML HTTP požadavky; data jsou přenášena ve formátu JSON. Autentizace uživatele je povinná a je delegována na server MUNI. Tato přístupnost dovoluje uživatelům pohodlný vstup z interaktivní osnovy jakéhokoli kurzu pomocí UČO a sekundárního hesla IS MUNI. Systém kromě testovaného subjektu nabízí tři další uživatelské role, které kopírují běžné využití v realitě a zachovávají větší míru zabezpečení položek. Celkem čtyři uživatelské role jsou: testovaný subjekt (student), kontakt na jednotlivá pracoviště s právem převádět výsledky studentů do IS (pomocník), tvůrce položek a testů (učitel) a správce. Systém tedy může sloužit nejen potřebám adaptivního rozřazovacího testu, ale učitelé mohou sami vytvářet testové baterie s konkrétní specializovanou slovní zásobou a položky si mohou nechat pretestovat dříve, než je zařadí do ostrých testů. Jen tak budou mít jistotu, že položky a z nich poskládané testy budou mít odpovídající úroveň podle SERRJ. Robustní subsystém rolí a oprávnění umožňuje spravovat objekty podobně jako v souborovém systému v Unixu (standard POSIX), a subsystém logování a reverze změn pro ochranu dat před zásahy většího počtu uživatelů.

Webové uživatelské rozhraní umožňuje kromě zadávání testů také uživatelsky příjemnou správu testových položek v položkové bance. Při návrhu byl kladen důraz na pružnost systému. Editor položek je přizpůsoben k efektivní práci s řády tisíců položek. Poskytuje pohodlné filtrování, tagování, označování a hromadnou editaci. Nedílnou

## Vznik a dosavadní vývoj adaptivního testu COMPACT



Obrázek 3.1: Vizualizace testantova průchodu

součástí je také automatické načítání položek při scrollování. Taktéž jsou ukládány jednotlivé verze položek pro revizi změn a sledování vývoje. Tagování bylo využito jako univerzální prostředek a slouží nejen po stránce popisné, ale i řídicí a organizační. Za zmínku dále stojí například vizualizace respondentova průchodu testem.

Uživatelské rozhraní z pohledu respondenta je přístupné z většiny moderních webových prohlížečů. Pro náročnější části aplikace jako editor položek a správa uživatelů doporučujeme používat prohlížeč Google Chrome nebo Mozilla Firefox. S prohlížeči rodiny IE máme špatné zkušenosti.

Adaptivní test je odzkoušen i zrakově postiženými studenty MU. Software pro předčítání nepočítá s dynamicky obnovovanou stránkou (XHR), pro zjištění nové informace musí být přečten znovu celý obsah. Pokud se však výsledek požadavku obalí do tagu A NAME a pod něj se umístí neviditelný hypertextový odkaz na tento tag, software už je schopen přečíst jen nový obsah.

Pro udržení funkčnosti a další rozvoj prostředí je nutné, aby budoucí vývojář systému měl velmi dobrou znalost jazyka JavaScript a jeho moderního užití. Nároky na znalost PHP a MySQL považuji za průměrné.

## Kapitola 4

# Testovací prostředí COMPACT

Tato kapitola se v první části věnuje způsobu, jakým se vybírají otázky z položkové banky do testu. V další části popisuje práci se systémem z hlediska studenta při průchodu testem. Třetí část je zaměřena na tvorbu a editaci položek, sestavování testů a jejich správu z hlediska učitele, a na převod výsledků do IS, což představuje hlavní náplň práce osoby v roli pomocníka.

### 4.1 Jak se vybírají otázky do rozřazovacího testu

Na rozdíl od testů, které si budou vytvářet sami učitelé, do rozřazovacího testu ve verzi 0.6 se podle dále popsaného modelu generují otázky z celé banky položek, které jsou označené znaménkem + (plus) a dále položky otagované jako pretest. Aby položka mohla být ozna-

## Testovací prostředí COMPACT

---

čena znaménkem plus, musí splňovat nastavené podmínky kalibrace. Každá položka může být v rámci jednoho průchodu testem položena nejvýše jednou.

Po dvou velmi jednoduchých, spíše vzorových otázkách se nejdříve pokládá stanovený počet položek k automatickému pretestu. Automatický pretest znamená, že systém náhodně vygeneruje položku z kterékoli skupiny označené znaménkem plus, jejíž informační hodnota z nějakého důvodu ve srovnání s ostatními nevykazuje prvotřídní parametry. To způsobuje, že systém s ní běžně nepracuje a položka stojí. Může to například nastat v případě, že položka se dlouhodobě nedržela nastaveného modelu a poměr jejího vyřešení, či nevyřešení se u testantů na odhadované úrovni výrazně vychýlil z poměru 50 na 50.

Další na řadě jsou čistě pretestové otázky (položky úplně nové nebo starší, ale opravené a vrácené do pretestovacího režimu), které se v určeném počtu náhodně volí ze skupiny označené štítkem jako pretest.

Výsledky klasické položkové analýzy, které se výše popsaným způsobem získávají, se v našem systému nevyjadřují v intervalu  $\langle 0, 1 \rangle$ , ale jsou namapovány na hodnoty odhadující úroveň respondenta v intervalu  $\langle -4, 4 \rangle$ . Je to proto, abychom se vyhnuli používání dvou různých měřítek a pro snadnější srovnání a zařazování položek zpět do ostrého testu. Výsledky z automatického ani klasického pretestu se studentům do jejich celkového výsledku nepočítají.

Naposledy přicházejí na řadu plně kalibrované položky, na jejichž základě se teprve stanoví výsledná úroveň testanta. Obtížnost prvních položek v této části testu je nastavena podle posledního nezrušeného výsledku z předchozího absolvování rozřazovacího testu. Pokud žádný takový není, protože testant vypracovává test poprvé, počáteční pozice se stanovuje na základě sebehodnocení z dotazníku. Z náhodně

## 4.2 Průvodce testem COMPACT

---

vybrané skupiny kalibrovaných položek, se vybere taková, která má obtížnost blízkou dosud vypočtené nebo odhadnuté úrovni respondenta. Pokud odpoví správně, jako další otázku dostane těžší, pokud odpoví chybně, následující otázka je lehčí. Postup se opakuje tak dlouho, dokud není vyčerpán definovaný počet položek. Test končí a výsledkem je úroveň vypočtená po poslední otázce.

Test máme nastavený ve dvou modelech. Jedním je diagnostický test, druhým je krátký placement. Na základě našich zkušeností stačí kolem dvaceti položek k ustálení výkonnostní křivky, a proto má krátký placement jen třicet položek, jejichž průchod se ještě urychle tím, že se studenta neptáme na jeho jistotu, s jakou na otázky odpovídal. Výsledkem je jen jeho odhadovaná úroveň vztažená k SERRJ.

Diagnostický test má kolem šedesáti položek, zjišťujeme při něm jistotu či nejistotu, s jakou testant jednotlivé položky zodpověděl, a výsledkem je kromě odhadované úrovně vztažené k SERRJ také summarizace gramaticko-lexikálních okruhů vyjádřených tagy položek, v nichž si byl jistý, ale odpověděl je chybně. K výsledku dále nabízíme radu, které učebnice by měl pro svůj další jazykový rozvoj používat nebo které volitelné předměty angličtiny na MU by mohl navštěvovat.

## 4.2 Průvodce testem COMPACT

Pro průměrně pokročilé uživatele se možná způsob, jakým popisujeme jednotlivé kroky od prvního přihlášení přes vytváření položek, jejich správu, nastavení a spuštění vlastního testu, analýzu pretestovaného materiálu podle klasické testové teorie s možností tvorby testů podle modelu IRT, až po zapsání výsledků studentům do IS, bude zdát zbytečně podrobný, ale naším cílem je, aby tento nástroj používali skutečně všichni vyučující CJV MU bez ohledu na počíta-

---

## Testovací prostředí COMPACT

---

čovou zdatnost. Jsme přesvědčeni, že správa položek i testů je natolik intuitivní, že to všichni bez problémů zvládnou, pokud se budou mít kam obrátit v případě jakýkoli nejasností. Při rozvoji testu počítáme s příručkou online, která se bude aktualizovat při každé změně a bude k dispozici u správce systému.

Na následujících stránkách se seznámíme s formátem adaptivního rozřazovacího a diagnostického testu tak, jak byl vytvořen pro splnění cílů projektu COMPACT. První část názorně ilustruje, jak studenti s testem pracovali a jaké zpětné vazby se jim po skončení dostalo.

Byla by škoda naprogramované prostředí a zkušenosti získané během projektu dále využívat jen k jednomu účelu – rozřazovacímu nebo diagnostickému testování pevně danému již vytvořenou položkovou bankou a modelem generování položek testantům odhadované úrovně – pokud tento nástroj nabízí i další využití. V další části se proto zaměříme na roli učitele. Popisujeme v ní jednotlivé kroky, které učitelé potřebují učinit k tomu, aby tento nástroj mohli používat pro potřeby testování ve svých jazykových předmětech. Tato část se zabývá vytvářením a správou položek a nastavením testů. Nakonec projdeme jednotlivé úkony, jichž je zapotřebí k tomu, aby byly výsledky testů převedeny do poznámkových bloků studentů v IS.

### 4.3 Průchod testem v roli studenta

Studenti se přihlásili na adrese adaptivního testu, kterou našli v interaktivní osnově do projektu zapojených předmětů, prostřednictvím svého UČO a sekundárního hesla do IS.

V prvním kroku museli studenti vyplnit krátký dotazník, který slouží jednak pro potřeby monitorovací zprávy projektu (v níž bylo důležité

### 4.3 Průchod testem v roli studenta

**Diagnostický test**

Moje testy Odhlásit

[Vytvořit nový test](#)

**Existující testy**

Před zahájením vyplňte, prosím, krátký vstupní dotazník. Tyto informace jsou povinné, protože jsou součástí monitorovací zprávy projektu.

Děkujeme.

**Státní příslušnost:**

**Pohlaví:**

**Vyberte možnost, která nejvíce vystihuje Vaše dovednosti v anglickém jazyce:**

Rozumím známým každodenním výrazům a zcela základním frázím, jejichž cílem je vyhovět konkrétním potřebám, a umím tyto výrazy a fráze používat. Umím představit sebe a ostatní a klást jednoduché otázky týkající se informací osobního rázu, např. o místě, kde žiji, o lidech, které znám, a věcech, které vlastním, a na podobné otázky umím odpovídat. Dokážu se jednoduchým způsobem domluvit, mluvím-li partner pomalu a jasně a je ochoten mi pomoci.

Rozumím větám a často používaným výrazům vztahujícím se k oblastem, které se mě bezprostředně týkají (např. základní informace o sobě a mé rodině, o nakupování, místopisu a zaměstnání). Dokážu komunikovat prostřednictvím jednoduchých a běžných úloh, jež vyžadují jednoduchou a přímou výměnu informací o známých a běžných skutečnostech. Umím jednoduchým způsobem popsat svou vlastní rodinu, bezprostřední okolí a záležitosti týkající se mých nejnaléhavějších potřeb.

Rozumím hlavním myšlenkám srozumitelné spisovné vstupní informace týkající se běžných témat, se

Obrázek 4.1: Úplně první přihlášení

dokládat, že se jedná o studenty s českou státní příslušností a uvádět počet mužů a žen), jednak jako sebehodnocení vlastní jazykové úrovně podle mírně upravených deskriptorů referenčního rámce tak, jak je známe například z Jazykového portfolia:

A1: Rozumím známým každodenním výrazům a zcela základním frázím, jejichž cílem je vyhovět konkrétním potřebám, a umím tyto výrazy a fráze používat. Umím představit sebe a ostatní a klást jednoduché otázky týkající se informací osobního rázu, např. o místě, kde žiji, o lidech, které znám, a věcech, které vlastním, a na podobné otázky umím odpovídat. Dokážu se jednoduchým způsobem domluvit, mluvím-li partner pomalu a jasně a je ochoten mi pomoci.

A2: Rozumím větám a často používaným výrazům vztahujícím se k oblastem, které se mě bezprostředně týkají (např. základní informace o sobě a mé rodině, o nakupování, místopisu a zaměstnání).

## Testovací prostředí COMPACT

---

Dokážu komunikovat prostřednictvím jednoduchých a běžných úloh, jež vyžadují jednoduchou a přímou výměnu informací o známých a běžných skutečnostech. Umím jednoduchým způsobem popsat svou vlastní rodinu, bezprostřední okolí a záležitosti týkající se mých nejnaléhavějších potřeb.

B1: Rozumím hlavním myšlenkám srozumitelné spisovné vstupní informace týkající se běžných témat, se kterými se pravidelně setkávám v práci, ve škole, ve volném čase atd. Umím si poradit s většinou situací, jež mohou nastat při cestování v oblasti, kde se tímto jazykem mluví. Umím napsat jednoduchý souvislý text na témata, která dobře znám nebo která mne osobně zajímají. Dokážu popsat své zážitky a události, sny, naděje a cíle a umím stručně vysvětlit a odůvodnit své názory a plány.

B2: Dokážu porozumět hlavním myšlenkám složitých textů týkajících se jak konkrétních, tak abstraktních témat včetně odborně zaměřených diskusí v mém oboru. Dokážu se účastnit rozhovoru natolik plynule a spontánně, že mohu vést běžný rozhovor s rodilými mluvčími, aniž by to pro ně představovalo zvýšené úsilí. Umím napsat srozumitelné podrobné texty na širokou škálu témat a vysvětlit své názorové stanovisko týkající se aktuálního problému s uvedením výhod a nevýhod různých možností.

C1: Rozumím širokému rejstříku náročných a dlouhých textů a rozpoznám implicitní významy textů. Umím se plynule a pohotově vyjadřovat bez výrazného hledání výrazů. Umím angličtinu užívat pružně a efektivně pro společenské, akademické a profesní účely. Umím vytvořit srozumitelné, dobře uspořádané, podrobné texty na složitá témata, čímž prokazují ovládnutí kompozičních útvarů, spojovacích výrazů a prostředků koheze.

C2: Snadno rozumím téměř všemu, co si vyslechnu nebo přečtu. Dokážu shrnout informace z různých mluvených a psaných zdrojů a při-



### 4.3 Průchod testem v roli studenta



#### Existující testy

Obrázek 4.2: Vzhled stránky po vyplnění vstupního dotazníku a při každém dalším přístupu

tom dokážu přednést polemiku a vysvětlení v logicky uspořádané podobě. Dokážu se spontánně, velmi plynule a přesně vyjadřovat a rozlišovat jemné významové odstíny dokonce i ve složitějších situacích.

Tento dotazník a sebehodnocení bylo třeba vyplnit pouze jednou při úplně prvním přihlášení se do systému bez ohledu na to, kolikrát student test během svého studia absolvoval. Výstup dotazníku nemá sice přímý náhled, ale správce je může pro další statistické zpracování exportovat z databáze.

Jedním z jeho hlavních posláních však je informace pro studenta a pro jeho další rozvoj jazykových kompetencí. Vzhledem k tomu, že deskriptory sebehodnocení odpovídají jednotlivým úrovním SERRJ, je možné studentům po absolvování testu sdělit rozdíl mezi tím, jak se sami hodnotí a jak je ohodnotil test.

Bez vyplnění těchto vstupních informací systém studenta dál nepustil. Teprve po řádném vyplnění se objeví základní stránka rozcestníku. Jde o stránku, kterou mají všichni účastníci společnou. Rozdíl mezi studentským přístupem a přístupem osoby s vyššími právy (učitele, pomocníka, správce) je v narůstajícím počtu nástrojů správy na tmavé liště mezi položkou Moje testy a Odhlásit.

## Testovací prostředí COMPACT

### Existující testy

04.01. 2012	<a href="#">Pokračovat v testu</a>
03.01. 2012	<b>118.1 bodů (C1 plus)</b> Test byl zrušen

Obrázek 4.3: Student má rozpracovaný test

Stránka *Moje testy* je vždy seznamem testů, který přihlášená osoba vyplnila, a je jedno, zda je to student, učitel nebo pomocník. Nejedná se tedy o „*Moje testy*“ ve smyslu autorství. Na obrázku 4.2 je znázorněna situace, kdy osoba zatím nemá absolvovaný ani rozpracovaný žádný test a má tudíž možnost vytvořit si nový test. Pokud by nějaký rozpracovaný test měla, objeví se pod existujícími testy nabídka *Pokračovat v testu*.

Po odkliknutí tlačítka *Vytvořit nový test*, nebo *Pokračovat v testu* v případě již rozpracovaného se na obrazovce objeví stránka obsahující obecné instrukce k danému testu. Dále pak student dostane na výběr, zda se chce seznámit s tutoriálem, či zda chce přejít přímo k testu.

Tutoriál s možností vyzkoušet si oba typy položek (každou navíc ve variantě s poslechem i bez), hraje pro studenty pozitivní roli jednak tím, že eliminuje první překvapení a navíc si studenti mohou svůj počítač ještě před spuštěním testu naostro zkontrolovat, zda jim technické nastavení přehrávače umožní bezproblémový poslech. V době, kdy nebylo možné test přerušit, byl výrazně vyšší počet studentů, kterým se musel test opětovně zpřístupnit, mezi těmi, kdo tutoriálem neprošli.

### 4.3 Průchod testem v roli studenta

---

Test můžete kdykoli přerušit zmáčknutím tlačítka "Po této otázce pozastavit test" a kdykoli se pak k němu vrátit.

Každá otázka, kromě poslechových otázek, má časový limit, který je dostatečně dlouhý na její zodpovězení. Po vypršení časového limitu je otázka automaticky odeslána s vyznačenou odpovědí. Pokud by vás ubíhající čas rušil, můžete jej skrýt.

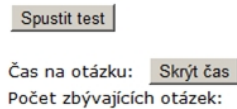
Nevyznačená odpověď a špatná odpověď jsou rovnocenné.

Doporučujeme projít si tutoriál pro práci s jednotlivými typy otázek.

Máte-li jakékoli další dotazy, nebo zaznamenáte-li během testování chybu, obraťte se prosím do [diskusního fóra](#) v ISu.

[Tutoriál](#) [Přejít k testu](#)

Obrázek 4.4: Instrukce před spuštěním testu



Obrázek 4.5: Spuštění testu

Velmi důležitým prvkem, jak pomoci studentům překonat izolovanost plynoucí z individuální práce online, je diskusní fórum, kam se mohou se svými problémy obracet. Osvědčilo se nám společné tematické diskusní fórum v IS vzhledem k tomu, že předmětů, v nichž se test používal, bylo hodně a problémy se řešily stále tytéž: přerušení internetového spojení a nevhodný prohlížeč patřily k těm nejčastějším.

Po kliknutí na tlačítko Přejít k testu dostávají studenti ještě jednu možnost se na test psychicky připravit a zároveň se rozhodnout, zda chtějí sledovat časování, nebo zda ho chtějí skrýt.

Kdykoli během testu se student může rozhodnout, že test po následující otázce pozastaví a vrátí se k jeho dokončení ve vhodnější chvíli. Kromě tohoto tlačítka se studentovi při každé otázce zobrazuje zbý-

## Testovací prostředí COMPACT

---

Po této otázce pozastavit test

Čas na otázku: **7** [Skrýt čas](#)

Počet zbývajících otázek: **61**

Obrázek 4.6: Informace k jednotlivým otázkám

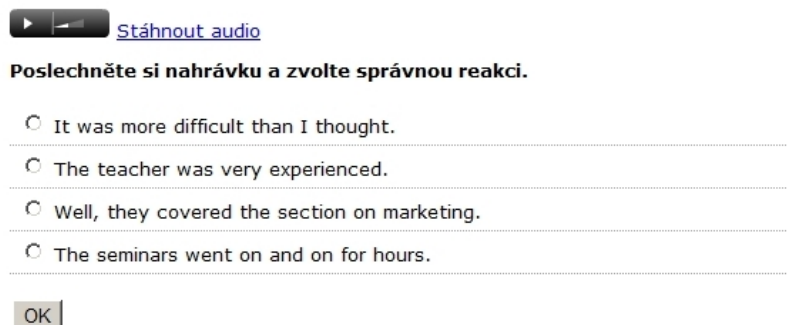
vající čas na právě zpracovávanou položku a počet všech zbývajících otázek do konce testu.

Jakmile se student definitivně rozhodne test spustit, vygeneruje se mu první otázka. Ve verzi rozřazovacího testu 0.6 jsme na začátek zařadili šest položek na porozumění slyšenému v pretestovacím režimu. U těchto položek nebyl důvod je časově omezovat, protože studenti jejich řešení na internetu ani jinde najít nemohli, ale přestože je mohli poslouchat tolikrát, kolikrát potřebovali bez omezení, nebyla to právě šťastná volba tím, že položky na poslech patří spíše k těm obtížnějším a řadu studentů vyděsily.

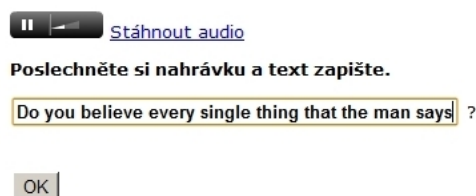
Otázky na poslech jsou v obou variantách – výběr z několika odpovědí i úlohy s tvořenou odpovědí. Kromě poslechových otázek, které byly zařazeny do podzimního testování poprvé, takže se pretestovaly úplně všechny, pretestovaly se další položky z kategorie tvořených odpovědí, které se nestihly pretestovat v jarním běhu, protože ten je pravidelně studenty méně obsazený, ale také mohlo jít o starší položky, u nichž došlo k natolik podstatné změně, že bylo nutno je znovu pretestovat. Studenti ale samozřejmě nepoznali, zda jde o otázku, která se do jejich celkového výsledku počítá, nebo je v testu pouze z důvodu pretestování. O vhodnosti tohoto způsobu průběžného zkvalitňování a rozšiřování položkové banky píšeme podrobněji v návodu k nastavování vlastních testů.

### 4.3 Průchod testem v roli studenta

---



Obrázek 4.7: Poslechová otázka výběru odpovědí



Obrázek 4.8: Poslechová otázka tvořená

Výše uvedené obrázky ilustrují oba typy poslechových otázek, s nimiž se studenti v rozřazovacím testu mohou setkat.

U výběru z několika odpovědí studenti volí správnou reakci na slyšený impuls ať už z textových nebo slyšených variant, tvořené položky u poslechových typů simulují diktát.

Po poslechových otázkách následovalo šedesát otázek, v nichž se různě střídaly položky na výběr z odpovědí s položkami tvořenými.

Mezi jednotlivými otázkami se student vždy musí rozhodnout, zda si byl řešením zcela jistý, nejistý, nebo zda nevěděl vůbec. O tom, že je

## Testovací prostředí COMPACT

---

Po každé otázce se objeví tento dialog. Odpověď neovlivňuje hodnocení, slouží ke sběru dat a Vám poskytne na konci testu zpětnou vazbu.

V předchozí otázce:

- si jsem jistý/á (77-100%),
- si nejsem jistý/á (34-76%),
- vůbec netuším (0-33%).

OK

Obrázek 4.9: Míra jistoty odpovědi na otázku

nutno se po každé odpovědi zamyslet, studenta informuje už tutoriál.

Tyto informace pak slouží k porovnání míry deklarované jistoty s realitou. Po vyhodnocení testu se student dozví, v jakém vztahu jsou jeho subjektivní pocity ohledně konkrétních gramaticko-lexikálních jevů ke skutečnosti. V zájmu udržení přehlednosti výstupních informací jsme považovali za zbytečné informovat studenta o tom, že neovládá něco, co on sám ví, že neovládá. Vygenerované hodnocení tedy vycházelo jen z gramaticko-lexikálních okruhů, u nichž si byl student svou odpovědí jist, ale položku přesto vyřešil chybně. Po absolvování testu se mu proto na obrazovce objevil text s hodnocením. To, že bychom pro informaci studentům zveřejňovali jednotlivé položky a jejich řešení, neplánujeme.

Například:

Právě jste absolvovali rozřazovací diagnostický test. Gratulujeme.

Dosažený počet bodů: 49

Odhadovaná úroveň podle Společného evropského referenčního rámce:  
B1

Vaše sebehodnocení: B2

U následujících gramaticko-lexikálních jevů jste si svou odpovědí byli jisti, ale udělali jste v nich chybu:

### 4.3 Průchod testem v roli studenta

---

- počitatelnost podstatných jmen
- použití členu neurčitého
- přací věty
- minulý čas prostý
- much/many/a lot/lots
- nepravidelná slovesa
- zájmena tázací
- členy v názvech států
- vazba „to be going to“
- vazba slovesa „look forward to“
- tvoření záporu u významových sloves
- tvoření přímé otázky
- rozdíl „must x mustn’t, need x needn’t“

Pro Vaše další studium angličtiny doporučujeme:

- učebnici: *English Grammar in Use: a self-study reference and practice book for intermediate students of English*, Raymond Murphy – Cambridge University Press
- volitelný předmět: *Online\_A*

Doporučujeme stažení a uložení této stránky do počítače.

Tím je studentův průchod testem skončený, výsledky uložené v databázi, kde čekají na svůj převod do IS, kterým se ale budeme zabývat až v podkapitole nazvané Administrace.

## 4.4 Jak se systémem pracovat v dalších rolích

V této části publikace se zaměříme na podrobný popis kroků, jichž je zapotřebí od vytvoření položky přes nastavení testu až po export výsledků testu do IS. Nejdříve se podíváme na stručný popis nových rolí, které bylo zapotřebí vytvořit, aby po skončení projektu mohlo s nástrojem, především pak s jeho vývojem, pracovat více lidí.

Další část se věnuje tvorbě a editaci vlastních položek, nastavení testů, administraci výsledků a jejich převodu do IS.

### 4.4.1 Stručný popis nových rolí

Systém nově rozlišuje čtyři role: student, pomocník, učitel a správce.

V roli studenta se ocitne každý, kdo je autentizovaný vůči IS a nemá přidělenou žádnou speciální roli. V definovaném období (od konce výuky v jednom semestru do předposledního dne konce změn v zápisu semestru následujícího) může po přihlášení jednou vypracovat diagnostický test, a pokud má od svého učitele odkaz na konkrétní test pro svůj jazykový předmět, může opět v učitelem definovaném období tento test jednou absolvovat. V případě jakýchkoli problémů se obrací do vyhrazeného diskusního fóra, kde získá potřebné informace. Při vstupu do systému osoba v roli student vidí lištu, na níž je pouze odkaz na Moje testy.

Náplní práce pomocníka je především převádět výsledky studentů do IS a být jim oporou při řešení problémů. Jeho činnost samozřejmě vyžaduje, aby měl v předmětu, do něhož se mají výsledky exportovat, i tomu odpovídající roli (musí mít právo zapisovat do poznámkových



#### 4.4 Jak se systémem pracovat v dalších rolích

---

bloků, exportovat a importovat z nich data). Má právo v případě objektivních důvodů studentův průchod zrušit, čímž mu umožní v definovaném období projít testem ještě jednou. Výsledky průchodů se nemažou, záznam o nich zůstane viditelný, ale uživatel si může nechat vytvořit další test.

Objektivním důvodem samozřejmě není, že student není spokojený se svým výsledkem, ale stát se může cokoli, co skutečně studentovi zabránilo test dokončit, včetně krátkodobého výpadku celého systému. Předchozí problémy se starými verzemi prohlížeče Explorer se už sice vyřešily, ale mohou nastat jiné nepředvídatelné situace. Při vstupu do systému osoba v roli pomocník vidí lištu, na níž je odkaz Moje testy, Export IS a Správa uživatelů.

Nejdůležitější rolí pro rozvoj tohoto testovacího prostředí je role učitele. Ta zahrnuje možnost:

- vytvářet vlastní položky;
- opatřovat je štítky (tagy) pro snazší orientaci při jejich vyhledávání;
- vybírat z databáze již hotových položek a buďto cizí položku použít beze změny, nebo ji zkopírovat a upravit;
- sestavovat z nich testy;
- připisovat k němu instrukce pro složení testu a hodnocení po skončení testu;
- zadávat časové období, kdy je možno test skládat.

Po vytvoření testu se vygeneruje webová adresa, kterou učitel například uvede do interaktivní osnovy předmětu v IS nebo jiným způsobem rozšíří mezi své studenty, kteří se jeho prostřednictvím dostanou

## Testovací prostředí COMPACT



Obrázek 4.10: Vstupní pohled učitele

k odpovídajícímu testu. Po vykonání testu pak studenti získají zpětnou vazbu, kterou si mohou z obrazovky běžným způsobem vytisknout nebo se k její uložené verzi v systému vracet. Učitel, stejně jako pomocník, může studentům průchody rušit, výsledky testu převádět do poznámkových bloků v IS a obsazovat roli pomocníka konkrétní osobou z MU definovanou číslem UČO.

Osoba v roli správce má práva dělat úplně všechno, ale vzhledem k tomu, že se jedná o pozici technickou, která má na starosti udržet funkčnost systému, bude se její činnost pravděpodobně omezovat na přidělování role učitelům a pomocníkům.

### 4.4.2 Pohled učitele

V této části si stručně popíšeme pohled učitele po přihlášení se do systému. Vysvětlíme funkce jednotlivých záložek, ale jejich podrobnějším popisem se budeme zabývat až v dalších částech této kapitoly spolu s ilustracemi, jak s nimi pracovat.

Po vstupu do systému se učiteli objeví odkaz *Moje testy*, ale zároveň na liště vidí i odkazy *Správa položek*, *Správa testů*, *Export IS*, *Správa uživatelů* a *Správa rolí*.

Odkaz:

- *Moje testy* je stejný pro všechny osoby v systému a zobrazuje všechny testy, které osoba absolvovala, tedy nikoli ty, jichž je

## 4.4 Jak se systémem pracovat v dalších rolích

---



Obrázek 4.11: Vyhledávání mezi položkami podle zvolených kritérií

autorem;

- Správa položek umožňuje vytváření a pozdější editaci položek včetně přidávání štítků (tagů) a prohlídku chování konkrétní položky;
- Správa testů slouží pro tvorbu a editaci testů;
- Export do IS umožňuje hromadný převod celých seznamů studentů s výsledky;
- Správa uživatelů pak k přístupu k průchodům jednotlivých studentů s možností tyto průchody rušit;
- Správa rolí dovoluje učitelům přidělit v systému roli pomocníka tím, že jeho UČO vepíše do pole formuláře.

### 4.4.3 Správa položek

Odkaz Správa položek přivede učitele na stránku, odkud může vytvářet nové položky a spravovat je, vyhledávat položky fulltextově nebo podle parametrů citlivosti, obtížnosti, parametru hádání nebo počtu, kolikrát se položka v testech již objevila.

Vpravo nahoře učitel najde čtyři odkazy (obr. 4.12):

## Testovací prostředí COMPACT

---

[Upravit tagy](#)  
[Vytvořit otázku](#)  
[Hřiště](#)  
[Přepočítat křivky](#)

---

Obrázek 4.12: Tvorba položek, editace tagů, hřiště a přepočet křivky

- úprava štítků (tagů),
- tvorba otázek,
- hřiště,
- přepočet křivky položky.

Jednotlivým odkazům a jejich funkcionalitám se věnujeme postupně v pořadí tak, jak jsou uvedeny na vstupní stránce správy položek.

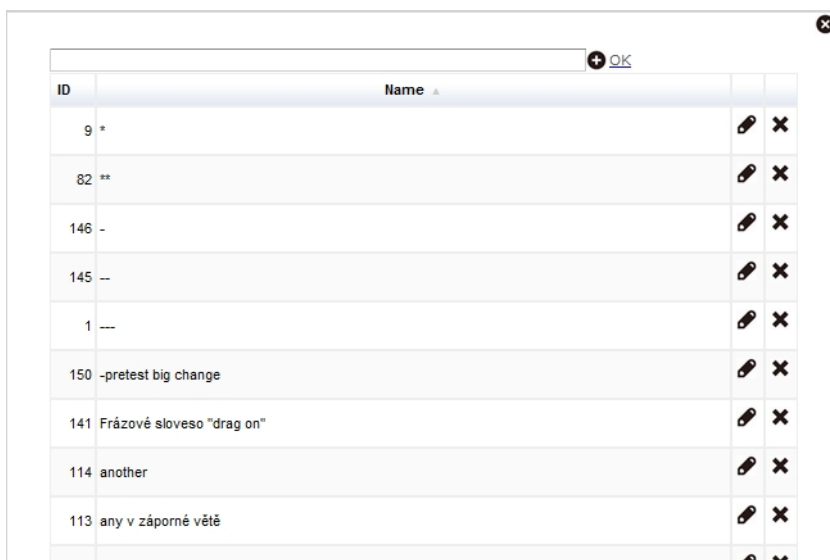
### Upravit tagy

Úprava tagů slouží k editaci již existujících štítků a k vytvoření nových. Štítky jsou velmi důležitým nástrojem jednak pro definování a přehlednost jednotlivých položek, jednak pro tvorbu testů. Při jejich vytváření buďte pečliví a snažte se vyhnout překlepům. Štítky lze samozřejmě kdykoli snadno opravit, ale hledání mezi překlepy je obtížné, zvláště pokud by v budoucnu bylo štítků velké množství. Tagy lze editovat pouze z této hlavní stránky správy položek, kdežto vytvářet je lze i v editačním režimu u každé položky.

### Vytvořit otázku

Tvorba nových položek probíhá tak, že po kliknutí na odkaz Vytvořit otázku zvolíme, který ze dvou typů chceme vytvořit. Volíme mezi po-

#### 4.4 Jak se systémem pracovat v dalších rolích



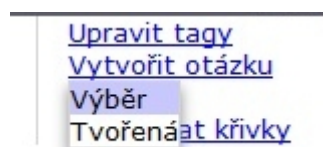
ID	Name		
9 *			
82 **			
146 -			
145 --			
1 ---			
150 -pretest big change			
141 Frázové sloveso "drag on"			
114 another			
113 any v záporné větě			

Obrázek 4.13: Přidávání nových a editace existujících tagů

ložkou typu výběr z několika odpovědí (pro stručnost Výběr) a typu otázka tvořená (Tvořená).

Po volbě typu otázky se dostáváme do jednoduchého formuláře. První formulář patří k otázce typu výběr odpovědí:

Vepíšeme kmen, tlačítkem přidat odpověď zvolíme celkový počet odpovědí (optimální je nejméně čtyři), vložíme a označíme jedno správné řešení, přidáme zvolený počet distraktorů, vybereme štítek a uložíme.



Obrázek 4.14: Typy položek

## Testovací prostředí COMPACT

The screenshot shows a web-based interface for creating a test question. At the top, there is a text input field containing the question "Can I park here?". Below this, there are four radio button options, each followed by a text input field and a close button (X):

- Only for half an hour.
- Sorry, I did that.
- It's the same place.
- I hope it was right.

Below the options is a button labeled "Přidat odpověď". Underneath that is a search bar containing the word "Modl". A dropdown menu is open, showing the following items:

- \*Modální slovesa
- modální slovesa "ought x should"
- Smazat statistiky

At the bottom of the interface are two buttons: "Uložit" and "Zrušit".

Obrázek 4.15: Formulář pro tvorbu položky typu výběr odpovědi

Do hlavního pole ještě před kmen testové otázky je někdy vhodné uvést instrukce, v našem případě by to mohlo být například: „K následující otázce zvolte nejlepší možnou reakci.“

Výběr štítků je velmi jednoduchý. Do vyznačeného pole píšeme podle svého uvážení nejvýstižnější název a z množiny tagů se dynamicky vybírají varianty existujících tagů zároveň s tím, jak je píšeme. Pokud se pro naše účely žádný z tagů přesně nehodí, jsme vyzváni k vytvoření nového.

Formulář pro tvořenou otázku je velmi podobný.

Jediný rozdíl ve formulářích mezi oběma typy položek je v tom, že do polí odpovědí vepisujeme všechna správná řešení, každé na jeden řádek. Podtržítkem vytvořená čára u tvořených odpovědí určuje, kde bude pro studenty vepisovací pole. S jeho délkou se nemusíte trápit, protože pole se dynamicky přizpůsobuje textu s tím, jak do něho

#### 4.4 Jak se systémem pracovat v dalších rolích

**4531**

<B><P>Doplňte správný slovesný tvar. </B></P>  
 She thought they would pay their share but unfortunately she \_\_\_\_\_ up paying for everybody.

ended	✘	(9)
finished	✘	(0)
wound	✘	(0)
landed	✘	(0)
fetched	✘	(0)

Přidat odpověď

\*\*\* | x \*Balanc pretest | x

[Nastavit audio](#)

Smazat statistiky

Uložit Zrušit

Obrázek 4.16: Formulář pro tvorbu tvořené otázky

student píše.

U tvořených otázek je u jakéhokoli počítačem podporovaného testování vždy problém, protože počítač uzná pouze tu správnou odpověď, kterou mu jako správnou vyznačíte bez ohledu na drobnosti, jakými je například více mezer mezi jednotlivými slovy. Absence tečky na konci věty nebo naopak o tečku navíc hodnotí počítač stejným způsobem, jako kdyby šlo o hrubku a naprosté nepochopení významu. Systém COMPACT nepočítá mezery a tečky navíc, nerozlišuje mezi českou čárkou a anglickým apostrofem, mezi českými a anglickými uvozovkami a nezajímají ho čárky ve větě. Přesto je běžné, že označí za chybu takové řešení tvořené odpovědi, které by učitel při opravě psaného textu na papíru za chybu nepovažoval.

Někdy je obtížné předjímat všechny možné správné odpovědi, které jsou studenti schopni vymyslet.

## Testovací prostředí COMPACT

---



Obrázek 4.17: Pole se prodlužuje podle délky textu automaticky

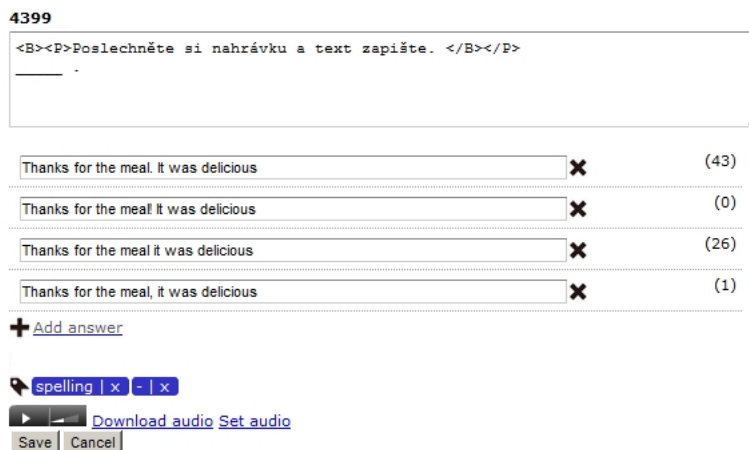
K tomuto účelu slouží mimo jiné právě i pretestování, během něhož se výsledky studentům sice nezapočítávají, ale sbírají se statistická data pro další osud položky. U každé tvořené otázky se všechny odpovědi ukládají a před nasazením otevřené položky do ostrého testu je nutné všechny tyto odpovědi projít a vybrat z nich případné další správné odpovědi.

Někdy učitel ví předem, že je více možných řešení (obr. 4.16), někdy je nutné, aby další řešení přidával na základě výsledků pretestování. Na obr. 4.18 je znázorněna situace, kdy z uložených odpovědí byly vybrány další, které sice přesně původnímu zadání neodpovídaly, ale které s ohledem na to, co jsme testovali (porozumění, pravopis) byly přijatelné.

Učitel se vždy musí rozhodnout, co vlastně testuje, a podle toho někdy zařadit jako správné řešení i nedokonalé odpovědi, protože zcela jiná je úroveň studenta, který na základě odposlechu vytvoří vlastní



## 4.4 Jak se systémem pracovat v dalších rolích



Obrázek 4.18: Další varianty uznané po pretestování jako správné

větu ve znění: „She looks like a famous filmstar“ a jinou úroveň naznačuje výsledná věta: „Shy looks like a fan this film star“ (obr. 4.19)

K editaci vytvořených položek se používá ikon, s jejichž pomocí se pak můžeme kdykoli po uložení k položce vrátit, opravit ji, smazat nebo udělat její kopii.

K oběma typům položek můžeme připojit zvukový záznam ve formátu mp3.

### Hřiště

Dalším odkazem na stránce vpravo nahoře ve správě položek je hřiště, které má sloužit k tomu, aby si učitel, který nemá s adaptivním testováním zkušenosti mohl vyzkoušet, jak se charakteristická křivka položky mění při různých parametrech a, b, c.

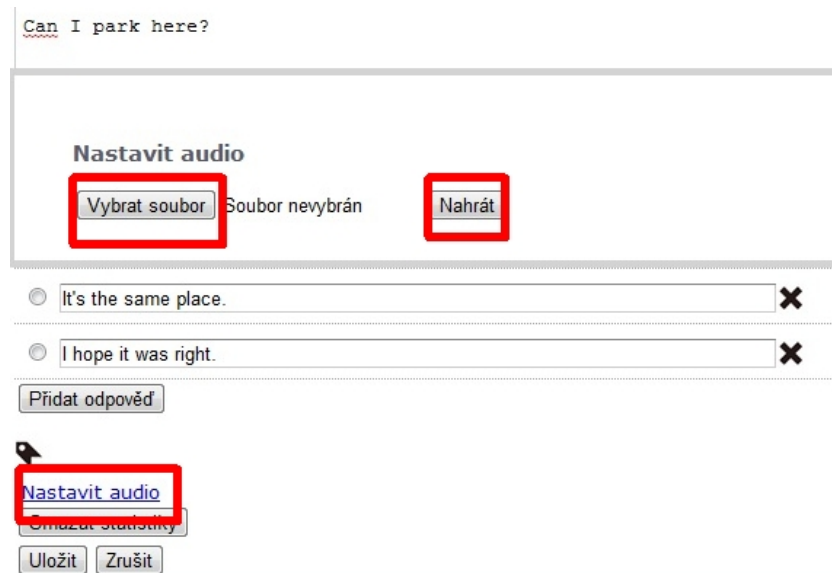
Zařadili jsme tento nástroj do systému proto, že o adaptivním testování a jeho modelech se toho mezi učiteli zatím mnoho neví. Poslá-

## Testovací prostředí COMPACT

Text	Count
She looks like a famous film star	233
She looks like a famous filmstar	52
She looks like a famous film star.	33
She looks like a famous filmstar.	11
She looks like a film star	7
She looks like a famouse film star	5
She looks like a fames film star	4
She looks like a famous films star	3
She looks like a famous film-star	3
She looks like a famous film's star	2
She looks like a film star.	2
She looks like a famous star	2
She looks like a fames films star	1
she looks like a fame filmstar	1
She looked like a famous film star	1
she looks likes a famous filmes	1
She lookes like a famous film star	1
She look like a famous filmstar	1
She looks like a famous film	1
She looks like a fame's film star	1
fiels like a famous film star	1
She looks like a films famous	1
She loks like famous film star	1
She looks like a famouse film stars	1
She luke	1
She looks like a fames film stars	1
She looks like a fmou film famous star	1
she looks like a famous filmsta	1
It's like a famous film star	1
She looks a like famous films star	1
She looks lake a famous film star	1
She looks like a famous movie star	1
She looks like a famous film stars.	1
She lokes like a famous film star.	1
Shy looks like a fan this film star	1
She looks like a fame film star	1

Obrázek 4.19: Různé varianty tvořených odpovědí

#### 4.4 Jak se systémem pracovat v dalších rolích



Obrázek 4.20: Vložení zvukového souboru

ním tohoto nástroje je, aby se učitelé seznámili s tím, jaký vliv mají změny v jednotlivých parametrech položky za následek ilustrovaný charakteristickou křivkou. A naopak - aby se naučili rozpoznávat, co jednotlivé křivky říkají.

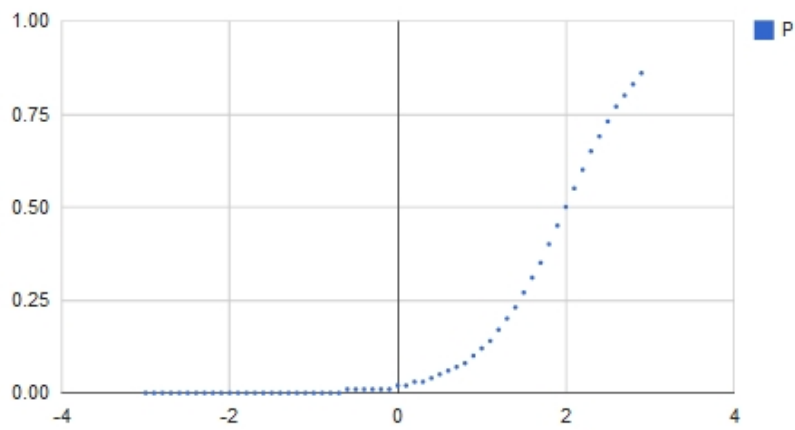
#### Přepočítat křivky

Posledním odkazem z nabídky ve správě položek přepočítání křivky. Charakteristické křivky položek se přepočítávají hromadně, ale je k tomu zapotřebí pokyn učitele. Systém sám kontroluje, zda se u položek nenasbíral dostatečný počet statistických dat, a učitel se jen na obrazovce při vstupu do správy položek objeví zpráva, že nastal vhodný čas k přepočtu. Kliknutím na odkaz systém přepočet provede.

## Testovací prostředí COMPACT

---

b: <-3, 3>   
a: <-1, 5>   
c: <0, 1>






Obrázek 4.21: Změny křivky v závislosti na různých hodnotách základních tří parametrů

## 4.4 Jak se systémem pracovat v dalších rolích

**1440** 25.09. 2011 a=2.59, b=0.05, c=0, d=0.003, počet: 306  
 We won't catch the train \_\_\_\_\_ home now! Please hurry up!

0 if we had left	(127)
0 until we are leaning	(22)
0 except we left	(27)
1 unless we leave	(129)
Timeout	(38)

podmínkové věty s podmínkou uskutečnitelnou \*\*\* \*Podmínkové věty  
[Zobrazit křivku](#)  
[Zobrazit změny](#)  
  

Obrázek 4.22: Parametry položky

### 4.4.4 Statistické informace o položkách

U položek jsou v jejich správě uvedeny parametry jejich kalibrace.

V záložce Správa položek je u každé otázky přehledně znázorněna řada důležitých informací.

Kromě identifikačního čísla položky a data poslední změny najdeme číselné vyjádření nejdůležitějších parametrů, které popisují informační hodnotu položky.

- $a$  = citlivost položky, informace o tom, zda a jak položka rozlišuje mezi dobrými a horšími studenty. Teoreticky může nabývat hodnot od mínus nekonečna do plus nekonečna, ale většinou se pohybuje mezi 0 a 2,8 (Baker 2001). Čím větší je tato hodnota, tím lépe úloha diskriminuje mezi testovanými. Položky s citlivostí dosahující záporné hodnoty by z položkové banky měly být vyřazeny hned po pretestování při vyhodnocení klasickou položkovou analýzou.
- $b$  = obtížnost položky může teoreticky opět nabývat hodnot od mínus nekonečna do plus nekonečna, ale většinou se její hodnota

## Testovací prostředí COMPACT

---

pohybuje mezi -3 a 3 (Baker 2001). Čím větší je hodnota  $b$ , tím je úloha obtížnější.

- $c$  = uhádnutelnost udává, jak velká je pravděpodobnost uhádnutí správné odpovědi. Jeho hodnota se teoreticky pohybuje mezi 0 a 1, v praxi většinou mezi 0 a 0,35 (Baker 2001).
- $d$  = odchylka od reality (čím je číslo menší, tím přesněji popisuje křivka vlastnosti otázky)
- počet vyjadřuje, kolikrát se položka objevila v pretestovém režimu, než byla zařazena do položkové banky

Čísla v závorce vedle správné odpovědi a jednotlivých distraktorů znamenají počet, kolikrát si je studenti zvolili. Tento ukazatel je také pro hodnocení položky důležitý, protože v případě, že se při pretestování zjistí nefunkčnost některého z distraktorů (nikdo si ho nezvolí), je třeba položku upravit a znovu pretestovat.

Další informace, jež lze z tabulky vyčíst, je počet studentů, kterým nestačil časový limit. Vypsány jsou také štítky a proklikem se lze dostat k předchozím verzím položky. U tvořených položek najdeme i seznam všech odpovědí s četností, s jakou se v testech vyskytly. K rychlému posouzení položky slouží charakteristická křivka položky, která jen ilustruje to, co vyplývá z ostatních parametrů: jedná se o průměrně obtížnou citlivě diskriminující položku.

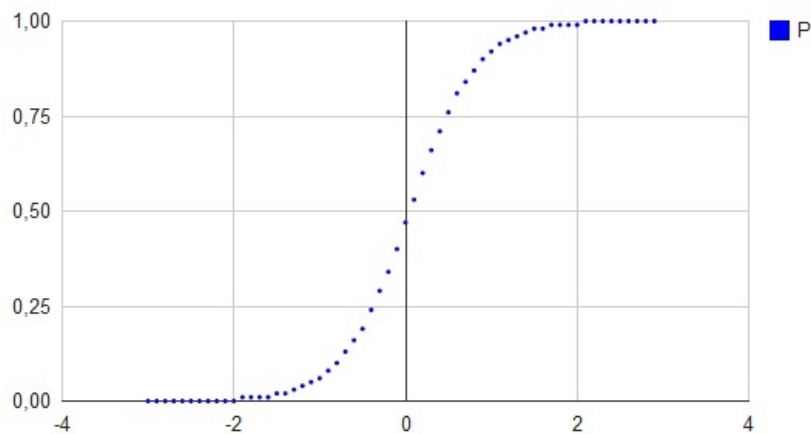
### 4.4.5 Nastavení vlastních testů

Nový test vytvoříme kliknutím na tlačítko Vytvořit test. Zobrazí se nastavení testu, ve kterém se zadává

- jméno testu;

## 4.4 Jak se systémem pracovat v dalších rolích

1440



Obrázek 4.23: Charakteristická křivka položky 1440

Moje testy Správa položek **Správa testů** Export IS Správa uživatelů Správa rolí

**Vlastní testy**

Název	Adresa pro přístup k testu			
adapt	<input type="text" value="http://adapt.lingua.muni.cz/test.php?test=1gz70u43h2ba"/>	<a href="#">Upravit termíny</a>	<a href="#">Upravit</a>	<a href="#">Smazat</a>

**Všechny testy**

Obrázek 4.24: První krok při tvorbě vlastních testů

## Testovací prostředí COMPACT

---

- instrukce ke zpracování;
- koeficient časového limitu otázky;
- počet položek k automatickému pretestování;
- a vytvářejí se skupiny tagů a známky.

### Vysvětlení některých pojmů

Skupiny tagů slouží ke specifikaci položek, ze kterých se bude test skládat. Z každé skupiny se při testu „náhodně“ vybírá daný počet otázek. Do skupiny tedy patří ty položky, které jsou označeny tagem, který je ve skupině uveden. Platí, že v rámci celého testu respondent nemůže dostat jednu položku dvakrát, není tedy třeba se bát zařazovat do skupin tagy, jejichž množiny položek nejsou vzájemně disjunktí. V nastavení skupiny se nastavuje počet položek, zda je skupina pretestová a přidávají se jednotlivé tagy. Pokud je skupina nastavena jako pretestová, odpovědi na položky neovlivňují výsledné hodnocení respondenta, ale slouží ke kalibraci položek.

Můžeme tedy vytvořit test sestávající se z jediné skupiny, která bude obsahovat veškeré tagy. V takovém případě nemáme zaručenou žádnou strukturu testu a všechny otázky mohou teoreticky být v extrémním případě položeny jen z jednoho gramatického jevu (prakticky je to téměř vyloučené).

Naopak můžeme vytvořit tolik skupin, kolik chceme do testu zařadit tagů, přičemž každá skupina obsahuje právě jeden tag. V takovém případě máme zajištěno, že se v testu objeví všechny jevy, které jsou danými tagy identifikovány. Ideální struktura testu se nachází nejspíše někde uprostřed, záleží ovšem na povaze testu.



## 4.4 Jak se systémem pracovat v dalších rolích

### Nastavení testu

#### Informace

**Název:** adapt

[Upravit](#)

**Instrukce:**

Dvakrát měř, jednou řež.

[Upravit](#)

**Diagnostický test:** Ne

[Upravit](#)

**Koeficient časového limitu otázky:** 1

[Upravit](#)

**Počet položek k automatickému pretestování:** 5

[Upravit](#)

#### Skupiny tagů

Pretest: Ne

Počet otázek: 10

názvy příbuzných a rodinných příslušníků

[Upravit](#)

[Odstranit](#)

[Přidat skupinu](#)

#### Známky

Název	Dolní hranice	Poznámka	-	-
Neprospěl			<a href="#">Upravit</a>	<a href="#">Smazat</a>
Prospěl	1		<a href="#">Upravit</a>	<a href="#">Smazat</a>

[Přidat známku](#)

Obrázek 4.25: Nastavení testů

## Testovací prostředí COMPACT

**Nastavení skupiny tagů**

Pretest:

Počet položek: 10 [Upravit](#)

[Přidat tag](#)

názvy příbuzných a rodinných příslušníků | x

[Upravit](#)

**Počet položek k automatickému pretestování: 5**

[Upravit](#)

**Koeficient časového limitu otázky: 1**

[Upravit](#)

**Skupiny tagů**

Pretest: Ne

Počet otázek: 10

názvy příbuzných a rodinných příslušníků

[Upravit](#) [Odstranit](#)

[Přidat skupinu](#)

**Známky**

Název	Dolní hranice	Poznámka	-	-
Neprospěl			<a href="#">Upravit</a>	<a href="#">Smaz</a>
Prospěl	1		<a href="#">Upravit</a>	<a href="#">Smaz</a>

[Přidat známku](#)

Obrázek 4.26: Nastavení skupiny tagů

#### 4.4 Jak se systémem pracovat v dalších rolích

učo

Instance ▾	Datum	Body	SERR	Zrušení	Pokračování	Zobrazení
11975	2011-12-31 09:14:01	52	B1	<a href="#">Zrušit</a>	Ne	<a href="#">Zobrazit</a>
11344	2011-10-03 10:50:34	47	B1	<a href="#">Zrušit</a>	Ne	<a href="#">Zobrazit</a>

Obrázek 4.27: Vypnutí časomíry

Některé položky v bance jsou v testu kvůli své nedostatečné kvalitě systémem ignorovány. K jejich dalšímu pretestování slouží automatický pretest. Ten respondentovi v testu předkládá určený počet nedostatečně pretestovaných položek, čímž je zaručeno, že se tyto položky, pokud budou mít vyhovující atributy (a, b, c, d), časem plnohodnotně zapojí do výsledného hodnocení respondenta), podrobněji viz následující závěrečná část této kapitoly.

Každá položka, kromě položek poslechových, má časový limit, který odrazuje od hledání odpovědi na internetu a který se vypočítává z typu otázky, délky zadání, počtu odpovědí a jejich délky (poslední dvě se týkají pouze výběrového typu). Výpočet limitu byl nastaven tak, aby respondentovi poskytoval více než dostatečný čas na zodpovězení otázky. Časový limit nesmí být omezující, protože by se jinak stal součástí obtížnosti otázky. Platí zde raději více nežli méně.

Nastavený koeficient se stane multiplifikátorem každého vypočteného časového limitu. Zůstane-li tedy koeficient roven jedné, vypočtený čas se nijak nezmění. Pro koeficient 0,5 se časový limit zkrátí na polovinu. Pro koeficient 2 se časový limit prodlouží na dvojnásobek. Chceme-li časový limit otázky pro celý test vypnout, nastavíme koeficient na nulu. Časový limit lze též vypnout jen pro konkrétní respondenty (viz Správa uživatelů).

## Testovací prostředí COMPACT

---

### Postup nastavení formuláře

Nastavíme jméno testu, instrukce k vypracování testu, které budou respondentovi zobrazeny před započítáním testu, a koeficient časového limitu otázky (doporučujeme ponechat na 1). Nastavíme počet položek k automatickému pretestu (ideálně alespoň pět položek).

Skupinu tagů vytvoříme kliknutím na tlačítko přidat skupinu. Zobrazí se nastavení skupiny, kde zaškrtneme zda skupina slouží k pretestování položek (v tom případě se odpověď na položku nepočítá do výsledného hodnocení respondenta). Uvedeme počet položek, který bude v testu z této skupiny respondentovi předložen. Nakonec přidáme jednotlivé tagy, ze kterých se mají položky v testu vybírat, kliknutím na tlačítko Přidat tag a následným výběrem tagu z nabídky. Tagy odebíráme podobně jako u otázek. Nastavení skupiny zavřeme křížkem v pravém horním rohu.

K výsledku testu lze přidat slovní hodnocení, nebo známku. Klikneme na Přidat známku. Zobrazí se nastavení známky, kde uvedeme její název (např. C1), dolní bodovou hranici pro udělení známky a popřípadě poznámku, která může obsahovat například doporučené učebnice, nebo jakoukoliv jinou informaci pro studenta. Nastavení potvrdíme tlačítkem Uložit.

Danou známkou bude ohodnocen respondent, jehož skóre je vyšší, nebo rovno, než daná dolní hranice a není vyšší než dolní hranice jiné známky. Pokud není dolní hranice uvedena, je známka považována za nejnižší. Respondentovi bude zobrazeno skóre, název známky a poznámka.

Nastavení testu zavřeme křížkem v pravém horním rohu.

Test je možné absolvovat jen pokud má aktivní termín.

V tabulce testů klikneme u příslušného testu na odkaz Upravit ter-

#### 4.4 Jak se systémem pracovat v dalších rolích

##### Nastavení známky

Název	<input type="text" value="Prospěl"/>
Dolní hranice	<input type="text" value="1"/>
Poznámka	<input type="text"/>
<input type="button" value="Uložit"/> <input type="button" value="Zrušit"/>	

Obrázek 4.28: Nastavení známky a jejích hranic

##### Nastavení termínů

Od	Do	-	-
2011-12-01 17:00:00	2012-03-02 23:59:59	<a href="#">Upravit</a>	<a href="#">Smazat</a>
<input type="button" value="Přidat termín"/>			

Obrázek 4.29: Nastavení termínů

míny. Zobrazí se nastavení termínů, kde klikneme na tlačítko Přidat termín. V nastavení termínu zadáme datum, od kdy je možné do testu vstupovat, a datum, kdy se test uzavře. Každý respondent smí test absolvovat v daném termínu nejvýše jednou. Jediná možnost, jak může respondent test v rámci jednoho termínu opakovat je zrušit jeho první výsledek ve správě uživatelů. Na základě testu a jeho termínu se též exportují výsledky (viz Export IS).

Ted' již zbývá jen zkopírovat odkaz na test z tabulky testů a vložit ho do osnovy předmětu, nebo rozeslat emailem studentům.

## Testovací prostředí COMPACT

---

Vyberte test:

Vyberte termín:

Nahrajte soubor z IS:  Soubor nevybrán

Obrázek 4.30: Nastavení termínu

### 4.4.6 Jak pretestovat

K nejkvalitnějšímu pretestování položek dochází při sestavení IRT testu z položek určených k pretestování a položek již dobře kalibrovaných, které jsou schopny úroveň respondenta přesně určit. Podle našich zkušeností na základě studia výkonostních křivek průchodu studentů během dvou semestrů je již počet kolem dvaceti kalibrovaných položek schopen podat poměrně přesný výsledek. Obě uvedené skupiny položek musí samozřejmě měřit stejný typ schopností respondenta (například obecnou angličtinu).

Pokud neexistují dostatečně kalibrované položky a je třeba začít v režimu klasického testování položkovou analýzou. V praxi to vypadá tak, že všechny položky, respektive skupiny v nastavení testu jsou označeny za pretestové. Je třeba se snažit, aby test byl co nejvyváženější. Jeho celková obtížnost by neměla být ani příliš vysoká, ani příliš nízká. Měl by sestávat z různě těžkých položek, jejichž zprůměrováním dostaneme průměrnou obtížnost. Totéž platí pro respondenty testu. Pokud tomu tak nebude, budou výsledné charakteristické křivky položek posunuty oproti realitě buď vlevo, nebo vpravo.

## 4.4 Jak se systémem pracovat v dalších rolích

Moje testy Správa položek Export IS Správa uživatelů Správa rolí

uČO

Uživatel nemůže vytvořit nový test.

Instance	Datum	Body	SERR	Zrušení	Pokračování	Zobrazení
12004	2012-01-04 00:09:42	-	-	<a href="#">Zrušit</a>	Ano	<a href="#">Zobrazit</a>
11987	2012-01-03 02:55:26	118.1	C1 plus	2012-01-04 00:09:04	Ne	<a href="#">Zobrazit</a>

Obrázek 4.31: Informace o jednotlivých studentech

### 4.4.7 Administrace výsledků a jejich převod do IS

Každý student vidí jen výsledky svých testů v záložce Moje testy. Výsledky jednotlivých studentů za každý jejich průchod testem vidí učitel nebo pomocník tak, že v záložce Správa uživatelů vepíše do pole UČO studenta.

Z této stránky může osoba v roli učitele nebo pomocníka prohlížet jednotlivé průchody a případně testy zrušit. I při zrušení testů zůstanou výsledky se všemi informacemi zachované pro přehled a další použití, ale zároveň zrušení testu umožní studentovi v definovaném období složit test ještě jednou.

Záznam na obrázku 4.31 vypovídá o tom, že student jednou testem prošel s výsledkem C1 plus, v současné době má rozpracovaný test, a pokud mu učitel tento průchod nezruší, v současné době student nový test vytvořit nemůže.

Pod tabulkou se zároveň s výsledky jednotlivých průchodů objeví přímý vstup do IS na osobní stránku konkrétního studenta s možností individuálního zápisu výsledků do zvoleného předmětu.

Častější než individuální přenos výsledků je ale přenos hromadný. Do poznámkových bloků v IS lze hromadně přenést výsledky prostřed-

## Testovací prostředí COMPACT

Uživatel nemůže vytvořit nový test.

Instance	Datum	Body	SERR	Zrušení	Pokračování	Zobrazení
6935	2011-07-21 22:21:42	-	-	2011-10-08 12:26:42	Ne	<a href="#">Zobrazit</a>
6913	2011-06-28 09:27:44	68.44	B2	<a href="#">Zrušit</a>	Ne	<a href="#">Zobrazit</a>
11944	2011-12-28 11:30:32	71.09	B2	<a href="#">Zrušit</a>	Ne	<a href="#">Zobrazit</a>
11528	2011-10-08 12:28:09	76.58	C1	<a href="#">Zrušit</a>	Ne	<a href="#">Zobrazit</a>

### Moje studentka

- [PdF\\_ONLINE\\_A](#) Angličtina Online
  - [Záznamník](#), [Indiv. informace o studentovi](#)
- [PdF\\_ZS1BP\\_AC2A](#) Cvičení z gramatiky on-line - AJ
  - [Záznamník](#), [Indiv. informace o studentovi](#)

### Výuka v minulosti

[Studium](#) [Profil](#)

### Studium

<b>Program</b>	<b>FF M-PS Psychologie, magisterský studijní program</b>
Forma	prezenční jednooborové
Stav	řádný zápis
<b>Obor</b>	<b>FF PS Psychologie</b>
Semestr	5. semestr, 1. cyklus
Forma	magisterské prezenční jednooborové

Obrázek 4.32: Vstup na osobní stránku studenta



#### 4.4 Jak se systémem pracovat v dalších rolích



Obrázek 4.33: Výběr poznámkového bloku v IS



Obrázek 4.34: Export seznamu studentů z poznámkového bloku

nictvím funkce export s následným importem.

Postup je jednoduchý a velmi rychlý v několika krocích.

Při prvním přenosu lze pracovat s celým seznamem studentů daného předmětu, při dalších převodech je rozumné omezit seznam studentů jen na ty, kdo v poznámkovém bloku zatím žádné výsledky nemají. K tomuto cíli nás dovede, když z nabídky Změna omezení zvolíme Dle textu v poznámkovém bloku, vybereme požadovaný blok a do pole pod seznamem bloků vepíšeme znak ! (vykřičník).

1. fáze probíhá v IS: vybereme poznámkový blok, kam chceme informace z adaptivního testu přenášet:

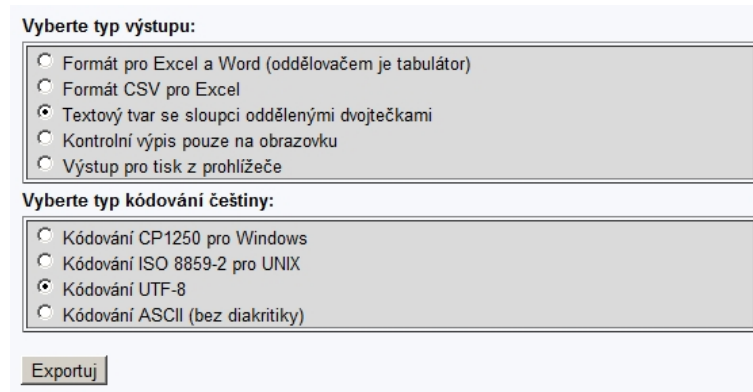
Jediné dvě operace, které nás pro převod výsledků po výběru bloku zajímají, jsou export a posléze import:

Odkazem export přistoupíme k výběru formátu a typu kódování:

Jedná se o textový formát se sloupci oddělenými dvojtečkami, kódování pak nastavíme na UTF-8.

Po stisknutí tlačítka Exportuj uložíme seznam studentů s názvem

## Testovací prostředí COMPACT



The screenshot shows a dialog box with two sections of radio button options. The first section, titled "Vyberte typ výstupu:", contains five options: "Formát pro Excel a Word (oddělovačem je tabulátor)", "Formát CSV pro Excel", "Textový tvar se sloupci oddělenými dvojtečkami" (which is selected), "Kontrolní výpis pouze na obrazovku", and "Výstup pro tisk z prohlížeče". The second section, titled "Vyberte typ kódování češtiny:", contains four options: "Kódování CP1250 pro Windows", "Kódování ISO 8859-2 pro UNIX", "Kódování UTF-8" (which is selected), and "Kódování ASCII (bez diakritiky)". Below these sections is a button labeled "Exportuj".

Obrázek 4.35: Výběr formátu výstupu

Moje testy Správa položek Export IS Správa uživatelů

Nahrejte soubor z IS:  No file chosen

Generuje pouze nové výsledky

Obrázek 4.36: Dialog pro generování výsledků testu z databáze

předmětu a poznámkového bloku do počítače.

1. fáze probíhá v testovacím prostředí COMPACT.

V záložce Export IS zvolíme tlačítko Chose File, v počítači vyhledáme soubor se seznamem studentů a odešleme. Vygenerovaný soubor se s dodatkem adapted opět uloží do stejného adresáře jako původní seznam, ale tentokrát už bude osazený výsledky.

2. fáze opět probíhá v IS. V záložce Poznámkové bloky tentokrát zvolíme operaci Import:

## 4.4 Jak se systémem pracovat v dalších rolích

---

Vybraný poznámkový blok ([vybrat jiný blok](#)):  
Final Test ONLINE\_A (469 stud.)

pořadí sloupce s identifikátorem studia (čís. od 1)  
 pořadí sloupce nebo sloupců s obsahem bloku (více hodnot odděluje mezerou)

Ignorovat změnu z původně exportovaného bloku na blok importovaný.

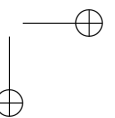
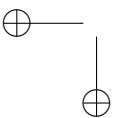
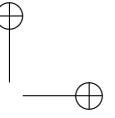
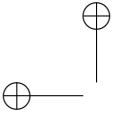
Zvolte režim:  
 import pouze na zkoušku, pro kontrolu chyb  
 import naostro

Vyberte soubor ve svém počítači:  
 No file chosen

Obrázek 4.37: Import

Vzhledem k tomu, že jde o odzkoušený proces, můžete přímo zvolit import naostro. Přes tlačítko Choose File najdete ve svém počítači soubor s doplňkem – adapted a tlačítkem Import ho uložíte do poznámkového bloku.

Možná to vypadá na první pohled složitě, ale jakmile si celou operaci od exportu přes generování nových výsledků až po import zautomatizujete, zjistíte, že je to záležitost trvající několik sekund.



## Kapitola 5

### Závěr

Exportem výsledků do IS jsme uzavřeli výčet dosavadních funkcionalit prostředí adaptivního testu COMPACT, ale na závěr se zkusme zamyslet nad tím, jakými cestami by se mohl ubírat jeho další vývoj.

Přestože pracoviště CJV MU vyučují převážně angličtinu, využívat tento nástroj by neměl být problém i pro další vyučované cizí jazyky.

Na MU ale studuje čím dál více zahraničních studentů i z jiných zemí než je Slovenská republika, kteří měli podle našich zkušeností při používání českého jazykového prostředí značné problémy, především proto, že navigace, instrukce a informace byly vytvořeno podle zadání projektu COMPACT pro české studenty. V další fázi vývoje tohoto nástroje by tedy stálo za zvážení vytvořit alespoň Anglickou jazykovou mutaci. Nejde o práci triviální vzhledem k pracnosti opatřit značný počet již existujících položek anglickými instrukcemi, ale při tvorbě nových položek by pak bylo jistě možné vkládat do položkové banky jednotlivé otázky s dvojjazyčným zadáním hned od začátku.

Práce na testovacím nástroji nekončí a neskončí ani se závěrem pro-

## Závěr

---

jektu. I nadále je třeba prostředí spravovat, aby nedopadlo jako diagnostický nástroj Dialang na serveru Lancaster University. Je třeba dále rozšiřovat položkovou banku i pro jiné testovací účely, sledovat chování položek, spravovat výsledky studentů.

V této souvislosti asi někteří učitelé namítnou, proč by to dělali, když mají k tomuto účelu, tedy k vytváření testů pro své studenty, odpovědníky v IS. Jedná se samozřejmě o pádnou námitku, ale tento projekt si klade za cíl využít síly spolupráce personálně silného pracoviště a společnými silami dotvořit komplexní nástroj pro další rozvoj standardizovaného testování, z něhož budou mít užitek studenti i učitelé napříč všemi fakultami MU. Pro jednoho učitele představuje tvorba standardizovaného adaptivního testu nepřekonatelný problém. Pro tým učitelů z celého CJV to už tak nedosažitelný cíl není, pokud je k dispozici nástroj, který spoustu výpočtů udělá za ně.

## Kapitola 6

### Literatura

- ALBERT, D., HOCKEMEYER, C., WESIĄK, G. (2002). Current Trends in e-Learning based on Knowledge Space Theory and Cognitive Psychology. *Psychologische Beiträge*, č. 44, str. 478-494.
- ALDERSON, J. Ch., HUHTA, A. (2005). The development of a suite of computer-based diagnosis tests based on the Common European Framework. *Language Testing*, 22 (3), str. 301 -302.
- ALTE Can Do statements and the CEFR. [Online]. Dostupné na <http://www.cambridgeesol.org/about/standards/can-do.html> v prosinci 2011.
- ANDRICH, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.
- BACHMAN, L. (1991). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

## Literatura

---

- BAKER, F. *The Basics of Item response Theory*. 2nd ed. [Online]. ERIC Clearinghouse on Assessment and Evaluation, 2001. ISBN 1-886047-03-0. Dostupné na: [www.ericae.net/irt/baker](http://www.ericae.net/irt/baker) v květnu 2009.
- BERNHARDT, E., & DEVILLE, C. (1991). Testing in Foreign Language Programs and Testing Programs in Foreign Language Departments: Reflections and Recommendations. In R. V. Teschner (Ed.) *Assessing foreign language proficiency of undergraduates* (str. 43-59). Boston: Heinle & Heinle.
- CANALE, M. (1986). The promise and threat of computerized adaptive assessment of reading comprehension. In C.W. Stansfield (Ed.), *Technology and language testing* (str. 29-45). Washington, DC: TESOL.
- CSIKSZENTMIHALYI, M. (1996). *O štěstí a smyslu života*. Praha: Nakladatelství Lidové noviny. ISBN 80-7106-139-5.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Dostupné na [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf) v lednu 2009.
- DENGLEROVÁ, D. (2007) *Alternativní přístupy k tvorbě a interpretaci psychologických testů*. Disertační práce. Brno: Masarykova univerzita, Psychologický ústav Filozofické fakulty. 94 s. Vedoucí dizertační práce doc. PhDr. Tomáš Urbánek, Ph.D.
- EMBRETSON, S. E, REISE, S. P. (2000). *Item Response Theory for Psychologist*. New Jersey: Lawrence Erlbaum Associates Publishers. ISBN 0805828192.



- 
- FILÍPKOVÁ, Z., BYČKOVSKÝ, P. (2008). *Studie proveditelnosti počítačem adaptovaného testování v prostředí českých škol*. [Online]. Dokument Studie\_CAT\_2008.pdf dostupný na [http://www.esfkvalita1.cz/Vystupy\\_projektu](http://www.esfkvalita1.cz/Vystupy_projektu) v prosinci 2011.
  - HAMBLETON, R. K., SWAMINATHAN, H., ROGERS, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: SAGE Publications, Inc. ISBN 0803936478.
  - HENNING, G. (1991). Validating an Item Bank in a Computer Assisted or Computer Adaptive Test: Using Item Response Theory for the Process of Validating CATS. *Computer Assisted Language Learning and Testing: Research Issues and Practice*. edited by P. Dunkel. New York, NY: Newbury House.
  - JELÍNEK, M., KVĚTON, P., DENGLEROVÁ, D. (2006). Adaptivní testování - základní pojmy a principy. *Československá psychologie*. 50, 2, str. 163-173. Praha: PSÚ AV ČR.
  - KAYA-CARTON, E. CARTON, A. P. DANDONOLI (1991). Developing a Computer Adaptive Test of French Reading Proficiency. *Computer Assisted Language Learning and Testing: Research Issues and Practice* edited by P. Dunkel. New York, NY: Newbury House.
  - LANGE, D. L. (1990). Priority Issues in the Assessment of Communicative Language Abilities. *Foreign Language Annals* 23 403-407.
  - LARSON, J. W. MADSEN, H. S. (1985). Computerized Adaptive Language Testing: Moving Beyond Computer-assisted Testing. *Calico Journal: The Computer Assisted Language Instruction*

## Literatura

---

*Consortium* [Online]. Dostupný z

<https://Calico.Org/A-252-computerized%20adaptive2011>.

- MEUNIER, L. E. (1994). Computer Adaptive Language Tests (Calt) Offer A Great Potential For Functional Testing: Yet, Why Don't They. *Calico Journal: The Computer Assisted Language Instruction Consortium*. [Online]. Dostupný na <https://Calico.Org/> v lednu 2011.
- MEN, H. (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling*. [Online]. Dostupné na <http://ir.uiowa.edu/etd/338> v červenci 2011.
- RASH, G. (1960). *Probabilistic Models for Some Intelligence Attainment Tests*. Chicago: University of Chicago Press.
- REICHOVÁ, H., KATRŇÁKOVÁ, H., HRADILOVÁ, A., ŠTĚPÁNEK, L., RYBIČKOVÁ, M., KUCHAROVÁ KALNÁ, Z., HULKOVÁ, H., ŠRÁMKOVÁ, E., MÉSZÁROSOVÁ, M., POJSLOVÁ, B., ROLLER BRESTŠÁK, I. *Popis jazykových kompetencí v odborném a akademickém jazyce - kritéria pro jejich posuzování*. Vyd. 1. Brno: Masarykova univerzita, 2010. 20 s. COMPACT, č. 1. ISBN 978-80-2105-114-0.
- SVOBODA, M. (2010). *Psychologická diagnostika dospělých*. Praha: Portál. ISBN 978-80-7367-706-0.
- TANG, K. L. (1996). *Polytomous Item Response Theory Models and Their Applications in Large-Scale Testing Programs: Review of Literature*. [Online]. Dostupné na: <http://www.ets.org/Media/Research/pdf/RM-96-08.pdf> v prosinci 2010

- 
- URBÁNEK, T., DENGLEROVÁ, D., ŠIRŮČEK, J. (2011). *Psychometrika*. Praha: Portál. ISBN 978-80-7367-836-4.
  - URBÁNEK, T., ŠIMEČEK, M. (2001). Teorie odpovědi na položku. *Československá psychologie*, ročník XLV, číslo 5, str. 428 - 440. Praha: PSÚ AV ČR.

**Adaptivní test COMPACT**

Autorský tým: PhDr. Tamara Váňová, Bc. Jiří Procházka,  
PhDr. Denisa Denglerová, Ph.D.

Vydala Masarykova univerzita v roce 2012

1.vydání, 2012, náklad 250 výtisků

Grafické zpracování a sazba: Bc. Jiří Procházka

Tiskárna Helbich, a. s., Valchařská 36, 614 00 Brno

ISBN 978-80-210-5742-5