

Fórum:Testy2/Zabezpečení testů

Dobré zabezpečení testů a testových otázek tvoří základní podmínku jejich věrohodného použití. Při neformálním průzkumu provedeném v roce 2007 mezi 30 000 americkými vysokoškoláky ^[1] se ukázalo, že 60,8 % jich během studia podvádělo. Tentýž průzkum ukázal, že 16,5 % z nich to nepociťuje jako etické provinění.

Poněkud překvapivé zjištění přinesl průzkum provedený na Fordhamově univerzitě: poukázal na významný rozdíl mezi studijními průměry podvádějících studentů a jejich poctivých protějšků. Podvodníci patří do skupiny se statisticky významně lepšími studijními výsledky než ti, kdo nepodvádějí. Je na místě klást si otázku, zda významnou roli při formování těchto postojů nehraje dnešní kultura, zaměřená na úspěch a nehlídící na druhé. V sázce je velké množství stipendií, ocenění, stáží a dalších pobídek. Je proto možné, že ti, kdo dosahují úspěchu nepoctivě, odůvodňují své chování právě odměňováním za „úspěch“.

Mají-li být výsledky testování věrohodným obrazem znalostí studentů, musejí být vyloučeny vlivy, které by výsledky mohly zkreslit.

Preventivní a metodická opatření

Součástí standardizace jsou preventivní a metodická opatření, která směřují mimo jiné k zajištění rovných podmínek pro všechny testované.

Před testem

V období před testem musí být zajištěn rovný přístup kandidátů ke znění testových otázek. Jsou dvě možnosti, jak riziko nerovného přístupu k otázkám snížit. Jednou je omezit exponování testu, tj. omezit příležitosti, při kterých může dojít k zaznamenání a přenesení testových položek do populace budoucích účastníků testu. Při každém použití testu toto riziko hrozí, i když ne vždy je naplněno. U testů s významným dopadem pro studenty je vhodné zveřejnit jen modelové otázky a položky použité v předchozích kolech testu považovat za pravděpodobně vynesené. Přístup k ostré verzi testu je třeba omezit na co nejužší okruh osob i v rámci pedagogického sboru, aby se snížilo riziko úniku položek ať již nedopatřením, či prolomením etických a profesionálních standardů.

Jiným způsobem regulace zmíněného rizika je navýšení počtu připravených otázek nad míru naučitelnou z paměti. Je-li v položkové bance několik tisíc položek, je ještě možné se odpovědi mechanicky naučit; pokud je však položek více, je již jednodušší se naučit látku samotnou. Regulace rizik se hojně diskutuje i v odborné literatuře ^[2].

Během testu

V době testu je třeba zajistit dozor, který vyloučí používání nepovolených pomůcek, opisování a napovídání. Některé školy v zahraničí zavedly režimová opatření týkající se odchodů studentů na toaletu v průběhu testování, neboť se ukázalo, že odchod zvyšuje riziko použití nedovolených pomůcek a napovídání. Při větším množství studentů je nezbytné zajistit i objektivní identifikaci účastníků testu, aby test nemohl složit někdo jiný.

Po testu

Vyplněné testy je třeba chránit před dodatečnými zásahy a zajistit jejich objektivní vyhodnocení. Jednou z cest ochrany je oddělení identity testovaného od zpracovávaného testu a jejich opětovné spojení až po kompletním zpracování a vyhodnocení testů.

Ani sebelepší preventivní opatření však nemohou podvádění zcela zamezit. Proto vznikly metody detekce podvádění na základě rozboru výsledků testu ^[3]. Krátce jsme se o nich zmínili již v kapitole 8 věnované analýze testů; některým se budeme věnovat v následujícím textu.

Následná detekce incidentů

Statistické vyhodnocení výsledků testů může pomoci identifikovat degradované položky, jejichž přínos pro objektivní a spravedlivé hodnocení byl narušen vynesením, popřípadě identifikovat účastníky testu, jejichž výsledky vykazují podezřelé znaky. Řada prací se zabývá přímo metodikou statistické detekce prozrazených otázek ^{[4], [5], [6]}. Přehled problematiky vyzrazených otázek ve vztahu k adaptivnímu testování podává souborná práce mapující aktivity na tomto poli v letech 1983 až 2005 ^[7].

Příkladem řešení problémů s testováním, které již proběhlo, byl zásah firmy Caveon specializované na bezpečnost testování. V lednu 2007 tato společnost řešila problém americké Federace státních komisí pro fyzikální terapii (FSBPT). Při detašovaných zkouškách fyzioterapeutů v Manile měla část účastníků testů k dispozici otázku, které zachytili účastníci předchozích testů; tyto otázky pak byly hromadně distribuovány. Přinutit všechny k zopakování testu by bylo obtížné – tíže důkazního břemene by se tím přenášela i na poctivé, navíc by hrozily žaloby na ušlý zisk za zpoždění licencí. Na druhou stranu potvrdit podezřelé výsledky testů by znamenalo ohrožit integritu celého testování a dobré jméno FSBPT.

V této situaci dostala společnost Caveon kompletní testová data ze všech testovacích míst za poslední dva roky k forenzní analýze. Důvodem byl předpoklad, že by zvýhodnění účastníci měli na otázky odpovídat neobvyklým způsobem. K identifikaci odlišně vyplněných testů použila společnost tři nezávislé statistické ukazatele, jejichž kombinací bylo možné dosáhnout pravděpodobnosti správné detekce s rizikem chyby menším než 1:1 000 000. Z prověřovaných 23 000 testů tak byla odhalena dvacítká, která měla všechny tři sledované ukazatele odchýlené od normálu. Na základě toho byly zmíněné testy prohlášeny za neplatné a ostatní uznány jako platné ^[8].

1. Online Education Database (OEDb). *8 Astonishing Stats on Academic Cheating* [online]. ©2010. Poslední revize 2010-12-19, [cit. 2012-11-25]. <<http://oedb.org/library/features/8-astonishing-stats-on-academic-cheating>>.
2. PITONIAK, Mary J. *item generation methodology in theory and practice*. Amherst : University of Massachusetts, School of Education, 2002. Center for Educational Assessment Research Report No. 468;
3. SHU, Zhan. *Detecting test cheating using a deterministic, gated item response theory mode* [online]. Greensboro : The University of North Carolina, 2010, dostupné také z <<http://libres.uncg.edu/ir/uncg/listing.aspx?id=4823>>. Vedoucí disertační práce Dr. Richard Luecht
4. YANG, Yongwei. *Exposed Items Detection in Personnel Selection Assessment: An Exploration of New Item Statistic* [online]. Annual meeting of the National Council of Measurement in Education in Chicago, ©2007. [cit. 2013-04-17]. <<http://www.measuredprogress.org/documents/10157/19213/ExposedItemDetection.pdf>>.
5. VEERKAMP, Wim.J.J a Cees.A.W GLAS. Detection of known items in adaptive testing with a statistical quality control method. *Journal of Educational and Behavioral Statistics*. -, roč. 25, vol. 4, s. 373-389, ISSN (Print) 1076-9986, (Online) 1935-1054.
6. HAN, Ning a Ronald HAMBELTON. *Detecting exposed test items in computer-based testing* [online]. Paper Presented at the Annual Meeting of the National Council on Measurement in Education, [cit. 2013-04-17]. <<http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/ha04-01.pdf>>.
7. GEORGIADOU, Elissavet G, Evangelos TRIANTAFILLOU a Anastasios A ECONOMIDES. A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment* [online]. 2007, roč. -, vol. 5, no. 8, s. -, dostupné také z <<http://escholarship.bc.edu/ojs/index.php/jtla/article/view/1647>>. ISSN 1540-2525.
8. FSBPT. *Forensic Analysis Conducted to Investigate Effect of Trafficking in Recalled Test Items Leads to Invalidation of 20 Candidate Test Scores* [online]. The Federation of State Boards of Physical Therapy, ©2012. [cit. 2012-12-13]. <<https://www.fsbpt.org/forfaculty/yourquestions/index.asp#InvalidatedNPTEScores>>.