

## ZKOUMÁNÍ VALIDITY ASSESSMENT CENTER

MAREK VRANKA

---

Intuitivně přitažlivá koncepce a relativně uspokojivá úspěšnost predikcí assessment center napomohla dlouhodobému přehlížení teoretických a metodologických problémů s nimi spojených. Až posledních několik let přineslo obnovení explicitního zájmu o tyto fundamentální otázky a možnosti jejich řešení. Záměrem článku je podat přehled o stávající situaci a výsledcích aktuálních výzkumů a napomoci tak jejich uplatnění v praxi.

**Klíčová slova:** assessment centrum, prediktivní validita, konstruktová validita

### Úvod

Assessment centrum (zkráceně AC) je často považováno za nejlepší existující metodu selekce zaměstnanců<sup>1</sup> a obvykle se takovéto její hodnocení přijímá jako daný fakt. Komplexní hodnocení chování v různých situacích skupinou odborných posuzovatelů je postup, který se pro odhad kvality uchazeče a predikci jeho budoucího pracovního výkonu zdá intuitivně vhodný nejen odborníkům ale i laické veřejnosti.<sup>2</sup> Zdálnivá neproblematičnost techniky AC tak vede k relativně malému zájmu o výsledky nejnovějších studií věnovaných jejímu kritickému zkoumání, a tedy i k pomalejšímu pronikání výzkumných závěrů do praxe.

V průběhu posledních pěti až deseti letech došlo k nárůstu odborného zájmu o techniku AC, především o její konstruktovou validitu a faktory ovlivňující její prediktivní validitu. Motivem pro napsání tohoto článku proto byla snaha sumarizovat znalosti o současném stavu problematiky AC a přispět tak k zavedení nejnovějších výzkumných poznatků do praxe.

Po úvodu následuje stručné vymezení základních pojmů a přehled vývoje zkoumání validity využití AC. Následující kapitola sumarizuje výsledky současných studií, jejichž implikace jsou v rámci diskuse podrobněji analyzovány v další kapitole. Text končí sumarizujícím závěrem.

---

<sup>1</sup> Teoreticky je AC možné použít pro výběr z uchazečů i v jiném než pracovním kontextu, pro zjednodušení ale v celém textu uvažujeme výběr zaměstnanců. S AC dále úzce souvisí i technika „development center“ (DC), její účel je však spíše diagnostický a proto jí nebude věnována bližší pozornost. Předmětem zájmu tohoto textu je AC v kontextu výběru nových a povyšování současných zaměstnanců.

<sup>2</sup> Bylo prokázáno pozitivní přijetí techniky AC ze strany posuzovaných, kteří ji považují za spravedlivou a objektivní (Macan et al., 1994; Smither et al., 1993).

## Základní teoretické pojmy

### Assessment centrum

Samotný pojem AC často není přesně vymezený. Pro účely tohoto textu je použita následující definice mezinárodní skupiny vydávající pravidla pro správnou organizaci AC: „AC představuje standardizovanou proceduru hodnocení chování založenou na datech z vícero zdrojů. K hodnocení je využito většího počtu zacvičených posuzovatelů a technik. Chování je posuzováno především na základě specificky vytvořených modelových situací. Jednotlivé posudky jsou následně sumarizovány buď při setkání posuzovatelů, nebo pomocí statistického procesu. V rámci integrační diskuse dochází k sumarizaci souhrnných záznamů chování – a často i sumarizaci jejich číselných hodnocení. Výsledkem této diskuse je hodnocení výkonu posuzovaného v dimenzích (nebo jiných proměnných), k jejichž měření bylo AC vytvořeno“ (International Task Force on Assessment Center Guidelines, 2009, s. 244–245).<sup>3</sup> Pokud technika nespĺňuje všechny uvedené body, případně v jakékoli své části využívá nevalidizované či předem neověřené postupy, nelze ji pokládat za AC (Vaculík, 2010).

### Validita

Pro metody selekce bývá tradičně za klíčovou pokládána prediktivní validita, tedy schopnost dané metody vybírat jedince, kteří budou pro výkon daných úkolů vhodnější, než jedinci vybraní náhodně. Tato validita je vyjádřena jako korelace mezi výsledky zjištěnými danou metodou a hodnotami cílového kritéria (hodnocení nadřizovaného, výše platu apod.) (Murphy, Shiarrella, 1997). Při výpočtu této korelace je nutné odstranit určitá zkreslení, která snižují její skutečnou hodnotu. Zejména se jedná o korekci nereliability v měření daného kritéria a korekci omezení rozsahu („range restriction“). V případě nereliability kritéria je výsledná hodnota prediktivní validity snižována nedostatky jiné metody (metody hodnocení pracovního výkonu) a reálná validita je proto vyšší (protože reálný výkon není ovlivněn nerealiabilitou v jeho měření) (van Iddekinge, Ployhart, 2008). „K omezení rozsahu dochází, pokud je ve validizačním vzorku variance prediktoru, kritéria, nebo obou menší než v relevantní populaci“<sup>4</sup> (van Iddekinge, Ployhart, 2008, s. 878). Taková situace nastává zpravidla u každé validizační studie, jelikož jsou posuzováni jedinci přijati/povýšeni na základě výsledků zkoumané metody a variance jejich výkonů je snížena o varianci odmítnutých jedinců. Skutečná validita je tedy vyšší než hodnota nekorigovaného koeficientu korelace.

Význam ostatních typů validity pro konkrétní praktickou aplikaci je řádově nižší. Obsahová validita může mít vliv na postoj vůči výběrovému řízení ze strany hodnocených účastníků (Murphy, 2009) a pro její stanovení obvykle postačuje expertní posouzení. Konstruktová validita se obvykle stanovuje pomocí metody „multitrait-multimethod“ (MTMM) matrice, při níž se srovnává korelace jednotlivých dimenzí v rámci jednoho

<sup>3</sup> Orig.: „An assessment center consists of a standardized evaluation of behavior based on multiple inputs. Several trained observers and techniques are used. Judgments about behavior are made, in major part, from specifically developed assessment simulations. These judgments are pooled in a meeting among the assessors or by a statistical integration process. In an integration discussion, comprehensive accounts of behavior – and often ratings of it – are pooled. The discussion results in evaluations of the assessee's performance on the dimensions or other variables that the assessment center is designed to measure.“

<sup>4</sup> Orig.: „RR occurs when there is less variance on the predictor, criterion, or both in the validation sample relative to the amount of variation on these measures in the relevant population.“

cvičení s korelací skóre dané dimenze z různých cvičení (pro každou dimenzi a každé cvičení) (Petrides et al., 2010). Tato validita zdánlivě (za předpokladu uspokojivé prediktivní validity) nemá pro personalistu žádný praktický význam.<sup>5</sup> Kritice takového názoru je věnována pozornost především v nejnovějších studiích (Arthur, Villado, 2008).

V současnosti při posuzování validity dominuje unitární přístup, podle něhož je nutná konvergence všech výše uvedených typů validity, aby bylo možné mluvit o celkové validitě použití metody pro daný účel za daných podmínek (Messick, 1995).

## Historický přehled zkoumání validity AC

Vznik prvních diagnostických nástrojů, které můžeme pokládat za předchůdce dnešních AC, se datuje do období II. světové války (Scroggins, Thomas, Morris, 2008). Z dobových publikací je patrné, že otázky metodologie, reliability a validity již v této době stály v centru pozornosti zainteresovaných odborníků. Ti při hodnocení prvních pokusů o AC sice uznávají obecný potenciál metody, s ohledem na její chybějící provázání s teorií a nedostatečně prokázanou validitu však zůstávají při celkovém posouzení AC spíše opatrně skeptičtí (Taft, 1959). Jejich výzvy k provedení řady pečlivých validizačních empirických studií a zlepšení teoretického ukotvení metody ale v průběhu 60. a 70. let utichají, spolu se vzrůstající vlnou popularity AC (Huck, 1973).

Na počátku tohoto nárůstu popularity stálo využití AC pro výběr zaměstnanců společnosti Bell System<sup>s</sup> v USA, jehož úspěšnost byla prokázána v empirické longitudinální studii (Bray, Grant, 1966). I přes určité metodologické nedostatky<sup>6</sup> byly výsledky této studie natolik povzbudivé (80 % úspěšných a 95 % neúspěšných manažerů bylo identifikovaných správně), že převládlo obecné přesvědčení, že pokud ne všechny, tak alespoň většina, metodologických obtíží spojených s AC byla úspěšně vyřešena. Ve skutečnosti je technika AC neproblematická pouze v otázkách reliability – postupy pro záznam hodnocení a trénink posuzovatelů jsou dobře zdokumentovány (Lievens, 2002; Reilly, Henry, Smither, 1990).

Co se týče validity, prvotní euforie se později ukázala jako předčasná. Z výsledků první metaanalytické studie výzkumů validity AC vyplynulo, že zatímco korelační koeficienty, zjištěné v rámci 50 studií, se pohybovaly v rozpětí od  $-0,25$  do  $0,78$ , hodnota korigované průměrné prediktivní validity byla pouze  $r = 0,37$  (Gaugler et al., 1987).<sup>7</sup> I když se tento údaj v následujících letech využíval jako zatím nejprůkaznější doklad validity AC, ve skutečnosti byla tato hodnota nižší, než se dříve odhadovalo. Definitivní zpochybnění postavení AC jako nejlepší metody selekce přinesla po 10 letech další metaanalytická studie srovnávající prediktivní validitu různých technik výběru zaměstnanců používaných v USA ve 20. století. Z ní vyplynulo, že nejvyšší prediktivní validitu při odhadování budoucí úspěšnosti přijímaného pracovníka má použití testů obecných rozumových schopností – prediktivní validita této metody je  $r = 0,51$ , což je hodnota podstatně vyšší, než byla zjištěna v případě AC (Schmidt, Hunter, 1998). Tento závěr podporují i výsledky novějšího metaanalytického výzkumu z evropského prostředí, ve kterém byla zjištěna prediktivní validita testu kognitivních schopností  $r = 0,56$  (Salgado, 2003).

<sup>5</sup> Uvedené neplatí v případě DC (Howard, 2008).

<sup>6</sup> Celkové hodnocení vycházelo kromě hodnocení výkonu ve cvičeních i z autobiografické skici, rozhovoru a asi desítky psychodagnostických testů tužka-papír a není tudíž možné stanovit přínos jednotlivých částí (Bray, Grant, 1966).

<sup>7</sup> Novější sedmiletý validizační výzkum v jisté holandské telekomunikační společnosti dospěl ke srovnatelné hodnotě validity ( $r = 0,39$ ) (Jansen, Stoop, 2001).

Odvodit z těchto výsledků závěr, že testy kognitivních schopností jsou v každé situaci pro výběr zaměstnanců vhodnější a AC by se nemělo vůbec používat, by bylo unáhlené a povrchní. Nicméně se AC musí s existující kritikou vyrovnat, a to pečlivým studiem faktorů ovlivňujících validitu, aby bylo možné určit v jakých situacích a jak přesně zkonstruované AC je pro výběr lepší než jiné metody.

## Přehled výsledků aktuálních studií

### Prediktivní validita jednotlivých dimenzí

Jedním z problémů AC je vysoký počet a variabilita sledovaných dimenzí chování.<sup>8</sup> Hodnocení účastníka v těchto dimenzích je následně obvykle zkombinováno do jednoho celkového skóre – tzv. OAR (overall assessment rating), které se využívá jednak pro rozhodnutí o přijetí uchazeče i ve validizačních studiích jako prediktor. Lze si klást otázku, jestli je sledování tak vysokého počtu různých dimenzí vůbec užitečné a jestli jsou hodnotitelé schopni bez využití statistických technik správně jednotlivá hodnocení integrovat.

Ve snaze zodpovědět tyto otázky Arthur s kolegy ve své metaanalýze vyabstrahovali 6 dimenzí běžně se vyskytujících v AC a zkoumali jejich prediktivní validitu ve srovnání s prediktivní validitou OAR ( $r = 0,37$ ) zjištěnou předchozím výzkumem (Arthur, Day, McNelly, 2003). Z jejich výpočtů vyplynulo, že dimenze „řešení problémů“ ( $r = 0,39$ ), „ovlivňování druhých“ ( $r = 0,38$ ) a „organizace a plánování“ ( $r = 0,37$ ) mají samy o sobě prediktivní validitu stejnou, nebo vyšší než OAR. Naopak dimenze „ohled na druhé“ ( $r = 0,25$ ), „drive“ ( $r = 0,31$ ) a „komunikace“ ( $r = 0,33$ ) dosáhly nižší hodnoty prediktivní validity než OAR (Arthur, Day, McNelly, 2003).

Autoři dále pokračovali vytvořením regresního modelu, jenž využívá jednotlivé dimenze jako prediktory, pomocí něhož se jim podařilo dosáhnout korelaci s kritériem  $r = 0,45$  (při vysvětlení 20 % variance), čím demonstrují nadřazenost statistické integrace před integrující diskusí. Jako nejsilnější prediktor se ukázala dimenze „řešení problémů“ (15 % variance), následně „ovlivňování druhých“ (+ 3 % vysvětlené variance) a „organizace a plánování“, resp. „komunikace“ (obě + 1 % vysvětlené variance). Zbývající dvě dimenze nepřinášely smysluplný informační přínos, a proto nebyly do výsledného regresního modelu zařazeny. Analýza také ukázala vysoké interkorelace mezi jednotlivými dimenzemi (průměrná odhadovaná interkorelace byla 0,56) (Arthur, Day, McNelly, 2003).

Z uvedené studie vyplývá, že zaměřením na hodnocení jediné dimenze („řešení problémů“) lze získat predikci budoucího pracovního výkonu minimálně stejně úspěšně jako při využití OAR sestávajícího z množství sledovaných dimenzí. Možnost zlepšit predikční sílu AC spočívá ve využití regresní analýzy, která přiřadí jednotlivým dimenzím (prediktorům) příslušné váhy a maximalizuje informační přínos pro rozhodování. Naopak jako kontraproduktivní se ukazuje sledování příliš velkého počtu dimenzí s vysokými hodnotami vzájemné korelace, což vede jen k vyšším nárokům na posuzovatele bez adekvátního zvýšení prediktivní validity.

<sup>8</sup> Polovina dotazovaných společností v USA i v Evropě ve svých AC sleduje 8 a více dimenzí, což je víc než doporučuje literatura (Eurich et al., 2009; Krause, Gebert, 2003).

## Konstruktová validita

Konstruktová validita, tj. zda hodnocený výkon v rámci určité části AC skutečně souvisí s měřenou dimenzí (a nikoliv s jinými dimenzemi), tradičně představuje slabé místo AC a výzkumy v této oblasti přinášejí smíšené výsledky již od 80. let (Hoffman et al., 2011). Obvyklým výsledkem je zjištění relativně vysoké korelace mezi různými dimenzemi v rámci jednoho cvičení a relativně nízké korelace mezi stejnými dimenzemi napříč cvičeními (Reilly, Henry, Smither, 1990). To znamená, že posuzovatelé hodnotí spíše než dimenze celkový výkon v jednotlivých cvičeních. Za možné příčiny tohoto jevu se považují nedostatečně operacionalizované dimenze, přílišná kognitivní zátěž pro posuzovatele, nebo odlišný potenciál cvičení pro pozorování všech dimenzí (tj. ne v každém cvičení je možné stejně dobře zaznamenávat chování vztahující se ke všem dimenzím) (Reilly, Henry, Smither, 1990).

Studie dokumentující nízkou konstruktovou validitu byly kritizovány jako zatížené metodologickými problémy a přítomností zkreslujících faktorů (jednalo se totiž často o výzkumy z praxe bez přísné metodologické kontroly) (Lievens, 2001b). V simulovaném AC, jež bylo navíc navrženo podle všech dostupných doporučení pro zvýšení konstruktové validity, posuzovatelé dovedli rozlišit jednotlivé dimenze, a zároveň ignorovat vliv cvičení, pokud byl výkon modelového uchazeče opravdu konstantní. Psychologové a studenti psychologie byli při rozlišování dimenzí úspěšnější než manažeři (Lievens, 2001b). V rozsáhlé studii analyzující MTMM matrice celé řady publikovaných článků pomocí konfirmační faktorové analýzy bylo následně demonstrováno, že konstruktová validita AC by mohla být lepší, než se dosud předpokládalo (Lievens, Conway, 2001). Tyto studie však byly později znovu analyzovány faktorovou analýzou s jiným modelem, jehož parametry lépe odpovídaly realitě. Nové výsledky závěry předchozí analýzy vyvrátily a potvrdily dominanci efektu cvičení (Lance et al., 2004). Na tomto případě můžeme zároveň vidět, že faktorové modelování činnosti AC představuje obtížný úkol, u něhož se nedá spoléhat pouze na matematickou korespondenci modelu a empirických dat. Pomocí metody Monte Carlo bylo prokázáno, že i „falešný“ model může zdánlivě dobře odpovídat datovému souboru (Lance, Woehr, Meade, 2007). Používaný model proto musí disponovat především kvalitními konceptuálními východisky.

Další studie přinášejí podobné závěry: konvergentní validita je sice nízká, ale vyšší v cvičeních, které umožňují lepší pozorování relevantního chování ( $r = 0,30$ ), než v cvičeních poskytujících jen omezený prostor k pozorování takového chování ( $r = 0,15$ ). Průměrná korelace dimenzí v rámci cvičení je přitom 0,50 (Haaland, Christiansen, 2002). Obdobné výsledky přináší i aktuální studie z Británie: nízká korelace jednotlivých dimenzí napříč cvičeními a vysoký efekt cvičení naznačují, že posuzovatelé nehodnotí dimenze, ale výkon v cvičeních (Petrides et al., 2010). Explorativní faktorová analýza navíc demonstrovala, že ve skutečnosti je celkové skóre určeno menším počtem faktorů, než je počet deklarovaných dimenzí, a tedy že dimenze mezi sebou silně korelují (Petrides et al., 2010). To naznačuje, že posuzovatelé se snaží rozlišit více vlastností, než ve skutečnosti dovedou, a údaj o sledovaných dimenzích by proto mohl být zavádějící.

Jedním z navrhovaných řešení této situace je úplné opuštění využití dimenzí a přechod ke konstruktům vymezeným podle jednotlivých cvičení (Lance, 2008). Teoretická východiska by se tedy podle tohoto přístupu měla přizpůsobit empiricky zjištěné faktorové struktuře celkového hodnocení. Vychází se při tom z tvrzení, že není důvod, aby byl výkon

posuzovaných účastníků stabilní v různých cvičeních, a že zaznamenaná konstruktová validita se zdá být problematická jenom na základě špatných předpokladů: MTMM matrice identifikují odlišné chování v odlišných situacích jako chybu, i když jsou takovéto odlišnosti vlastně smysluplné a očekávatelné. Ve skutečnosti má smysl hodnotit jen jednotlivá cvičení a z těchto hodnocení syntetizovat celkové skóre jako tzv. obecný faktor výkonu v AC (Lance, 2008).

Navrhovaný postup však byl spíše kritizován jako unáhlený a nepsychologický (Arthur, Day, Woehr, 2008; Howard, 2008). Na jedné straně je pravda, že cvičení v AC nikdy neměla měřit všechny dimenze stejným způsobem, naopak měla poskytovat odlišné úhly pohledu na dané dimenze. Zároveň však nemůžeme říct, že se v různých cvičeních nesledují vůbec žádné společné dimenze (Howard, 2008).

Alternativním vysvětlením nízké konstruktové validity podporovaným větší skupinou odborníků je, že přestože se v rámci AC deklaruje měření určitých dimenzí a konstruktů, tyto konstrukty se ve skutečnosti neměří (Arthur, Day, Woehr, 2008). Toto vysvětlení podporuje i skutečnost, že v cca 40 AC bylo sumárně identifikováno více než 150 různých (údajně) sledovaných dimenzí, přičemž je těžko představitelné, že by jich i přes komplexní povahu manažerské činnosti bylo k jejímu vysvětlení nutné takové množství (Arthur, Day, Woehr, 2008). Na druhé straně by dimenze neměly být příliš vágní a mít povahu osobnostních rysů, jelikož odhadovat na základě chování osobnostní rysy vyžaduje značnou míru spekulace (Howard, 2008). V tomto ohledu se jako nejvhodnější počet jeví šest obecných dimenzí identifikovaných ve výzkumu Arthura a kolegů, jenž byl popsán výše (Arthur, Day, McNelly, 2003). Podle tohoto vysvětlení by nízká diskriminační validita zjištěná v důsledku vysoké interkorelace dimenzí byla zčásti způsobena existencí obecných dimenzí, sytících primární dimenze.

Kromě ledabylého přístupu k tvorbě (údajně) sledovaných dimenzí je kritizována i metodika MTMM matric, protože pracuje s hodnocením dimenzí po jednotlivých cvičeních (tzv. PEDR – „post exercise dimension rating“), i když pro výsledné hodnocení má větší význam posouzení dimenzí po administrování všech cvičení (tzv. PCDR – „post consensus dimension rating“). V důsledku této praxe může docházet k inflaci efektu cvičení (Arthur, Day, Woehr, 2008).

### **Nový model AC a jeho ověření**

Celkem lze popsat čtyři základní modely v minulosti využívané pro zkoumání konstruktové validity AC, přičemž první dva jsou spíše hypotetické: (a) *J*-dimenzí 0-cvičení a (b) 0-dimenzí *K*-cvičení, podle nichž by měl být v rámci AC pozorován výhradně vliv dimenzí (model a), resp. vliv cvičení (model b). První model by byl ideálem z psychometrického hlediska, zatímco druhý si lze představit jen jako sérii nesouvisejících, vysoce specifických úkolů. Další dva modely lépe odpovídají empirickým nálezům: (c) *J*-dimenzí *K*-cvičení, podle něhož existuje jak efekt cvičení, tak efekt dimenzí, a (d) 1-dimenze *K*-cvičení, který odpovídá Lanceovu návrhu a kromě vlivu cvičení předpokládá i faktor „obecného výkonu v AC“. Na základě výše shrnutých poznatků byl navržen pátý model, jenž by měl nejlépe vystihovat skutečný mechanismus posuzování v AC: (e) *J*-obecných dimenzí *K*-cvičení 1-faktor obecného výkonu (Hoffman et al., 2011). Obecné dimenze (viz Arthur, Day, McNelly, 2003) vycházejí z předpokladu, že část společné korelace mezi původními sledovanými dimenzemi není způsobena chybou, ale že naopak dokazuje existenci obecnější dimenze, jež je společná několika původním.

Tento model byl testován v rozsáhlém výzkumu na čtyřech AC ( $N$  od 88 do 1075), přičemž pro lepší poznání konstruktové struktury AC bylo využito i srovnání s výsledky externích instrumentů. Konkrétně se jednalo o faktory extraverte a svědomitosti měřené pomocí osobnostního dotazníku a inteligenci zjištěnou pomocí testu obecných rozumových schopností. Relevance zmiňovaných faktorů pro výkon povolání byla prokázána celou řadou jiných výzkumů (viz Barrick, Mount, 1991; Salgado, 2003). Výsledky konfirmační faktorové analýzy prokázaly, že poslední model (resp. všechny jeho varianty lišící se mírně odlišnou konstrukcí obecných dimenzí) odpovídá empirickým datům nejlépe ze všech modelů (Hoffman et al., 2011). Tento výsledek pochopitelně nelze vnímat jako nezvratný důkaz, nicméně představuje silný argument ve prospěch platnosti navrhovaného modelu, kromě jiného i proto, že ošetřuje řadu námitek v minulosti vznášených proti výsledkům konfirmační faktorové analýzy (viz Lance, Woehr, Meade, 2007). V průměru bylo možné vysvětlit 27 % variance v celkovém hodnocení pomocí hodnocení v jednotlivých dimenzích, zatímco pomocí hodnocení v jednotlivých cvičeních jí bylo možné vysvětlit až 41 %, což je v souladu se staršími výzkumy a zároveň představuje argument proti návrhu na úplné vyloučení posuzování dimenzí z AC (Hoffman et al., 2011).

Složení obecných dimenzí se v jednotlivých zkoumaných AC lišilo, obvykle ale zahrnovaly „interpersonální“ a „koncepční/administrativní schopnosti“, „vůdčovství“ a dále pak „komunikaci“ a „drive“ (Hoffman et al., 2011). To dokládá, že těchto obecných dimenzí je relativně nízký počet a zároveň jsou možné jejich kombinace podle nároků konkrétní pracovní činnosti (např. přiřazením různých vah při konečném hodnocení). Samozřejmě lze namítnout, že na základě stručného popisu údajně sledované dimenze není možné spolehlivě určit, co se doopravdy při hodnocení této dimenze měří (Arthur, Day, Woehr, 2008). To ale nebylo cílem Hoffmanové analýzy, a navíc, jelikož se i přes uvedené limitace povedlo získat smysluplné výsledky v souladu s očekáváními podloženými literaturou, svědčí to o jisté důvěryhodnosti takto založené klasifikace dimenzí.

Celkově je možné říci, že nejnovější model, sestávající z obecných dimenzí, cvičení a faktoru obecného výkonu, popisuje fungování AC nejlépe ze všech dosud navrhovaných modelů jak z hlediska teoretické koncepce, tak svou korespondencí s empirickými daty. Implikace tohoto zjištění jsou následující: (a) Hodnocení dimenzí poskytuje významný informační přínos pro celkové hodnocení a jejich využívání by se v současnosti rozhodně nemělo opouštět. (b) Je potřebné prohloubit teoretické poznání dimenzí, jejich příslušnosti k obecným dimenzím a o vztahu cvičení a jednotlivých dimenzí (Lievens, 2008). (c) Příslušné konstrukty je nutné identifikovat i v rámci analýzy pracovní činnosti a analýzy pracovního výkonu a úspěšnosti.

## Diskuse

Nedostatečně rozvinutá teorie osobnosti a kognitivní teorie v minulosti neposkytovaly prostředky pro snadnou identifikaci prakticky využitelných konstruktů, což mohlo být jedním z důvodů, proč se personalisté a teoretici AC uchýlili k substituci konstruktové validity za obsahovou validitu. Vycházelo se při tom z myšlenky, že pokud je činnost posuzovaná v daných cvičeních co možná nejpodobnější činnosti reálně vyžadované při výkonu povolání, budou manažeři schopni díky svým tacitním znalostem rozpoznat nejuhodnější kandidáty. Sledovaná činnost by tak nebyla „znakem“ odkazujícím na budoucí výkon, ale

jednoduše „ukázkou“ skutečného výkonu (Chen, 2006). Tím se na první pohled elegantně odstranil problém s konstrukty, na druhé straně však přestalo být AC vnitřně konzistentní. Pokud se totiž opustí explicitní využití konstruktů, ztrácí smysl i úvodní analýza pracovní činnosti a hlavně hodnocení jednotlivých dimenzí, které ale představuje nutnou součást AC (viz International Task Force on Assessment Center Guidelines, 2009). Interpretace a integrace hodnocení jednotlivých dimenzí do celkového hodnocení (OAR) tím navíc ztrácí jakoukoliv teoretickou oporu, protože ve skutečnosti se hodnotí spíše cvičení než dimenze. Jelikož však personalistům postačovalo, že byli s pomocí OAR schopni relativně úspěšných predikcí, otázka konstruktů se octl mimo centrum pozornosti.

Situace se začala měnit až v posledních pěti, nanejvýš deseti letech, kdy se charakter AC jako „černé skřínky“, pravděpodobně fungující ve značném rozporu se svým deklarovaným mechanismem, ukázal jako neudržitelný. Jedním z navrhovaných řešení pro odstranění tohoto rozporu bylo přizpůsobit teoretická východiska výše popsané praxi (Lance, 2008). Takovýto přístup by vedl spíše k tzv. „work-sample“ technice, tj. metodě selekce, u níž se posuzuje uchazečův výkon při provádění úkonu tvořícího významnou či běžnou součást daného povolání (Bucalo, 1974). Technika se využívá spíše u manuálních povolání a hodnotí se především správnost (případně rychlost, bezchybnost, kvalita) provedení. Neexistuje zásadní důvod, kvůli němuž by se podobný přístup nemohl použít i při výběru manažerů či jiných nemanuálních zaměstnanců. V takovém případě je však místo současných dimenzí nutné identifikovat (na základě vhodně přizpůsobené analýzy pracovní činnosti) kritéria hodnocení správnosti daného výkonu. Konstruktům by se tedy nešlo vyhnout úplně, změnila by se jen základní koncepce jejich tvorby. Je pravda, že využitím vysoce specifických úkolů, pokrývajících většinu pracovní náplně, by bylo možné ztotožnit úkol s konstruktem. Čím však bude využit specifitější úkol, tím omezenější informaci získáme: jedinec může dosahovat dobré výsledky v dané činnosti díky minulým zkušenostem a může například selhat po změně pracovních požadavků. Otázkou také zůstává, do jaké míry by se takováto technika mohla i nadále nazývat AC, a jestli by spíše nešlo o zánik AC jako takového.

Většina odborníků zabývajících se danou problematikou se však postavila proti takovému postupu a za zachování původních východisek techniky AC. Tento návrh naopak obnáší podrobnější zkoumání a případnou úpravu praxe podle paralelně se rozvíjejícího teoretického rámce. Jedná se zejména o bližší prozkoumání struktury dimenzí a obecných dimenzí relevantních ve vztahu k pracovnímu výkonu (přičemž postup by mohl být částečně analogický sestavení BIG5 modelu v teorii osobnosti), ale také o zkoumání potenciálu určitých typů cvičení vyvolávat chování, na základě něhož lze dané dimenze hodnotit. A v neposlední řadě jde o poznání míry specifity/zobecnitelnosti chování, pozorovaného v různých typech cvičení. Pokud by se na podobných principech analyzovala i samotná pracovní činnost, bylo by pravděpodobně možné stanovit velmi silné a teoreticky dobře podložené spojení mezi dimenzemi a výkonem.

Poslední studie naznačují, že k tomu povede ještě dlouhá cesta. Pro praxi lze prozatím doporučit používání metod zvyšujících diskriminační validitu mezi jednotlivými dimenzemi, tj. např.: již zmiňovaná redukce počtu dimenzí, zaškrťovací seznamy, hodnocení v rámci dimenzí, ne v rámci jednotlivých úkolů (Arthur, Day, McNelly, 2003) a využití odborníků v HR a psychologů jako posuzovatelů místo manažerů (Lievens, 2001a).



## Závěr

Psychodiagnostická koncepce, z níž praxe AC před téměř 70 lety vzešla, byla svého času vrcholem ve snaze o vědecké zkoumání lidského jedince a predikci jeho budoucího chování. To se ale postupem času měnilo; zatímco studie z 50. let důrazně připomínají limitace a nejrůznější problematické aspekty metody, jakož i existenci smíšených výsledků z její aplikace v klinickém prostředí (Taft, 1959), postupem času začínají v publikované literatuře převládat pouze pozitivní hodnocení. Vytváří se tak dojem, jako kdyby všechny v minulosti se objevivší problémy byly úspěšně vyřešeny, i když ve skutečnosti se jim již jen nevěnuje pozornost. Analýza pracovní činnosti a stanovování sledovaných dimenzí se dnes v praxi často považují za neproblematické, i když původně představovaly jednu z hlavních metodologických obtíží: jak určit, které charakteristiky jsou opravdu důležité, a jakým způsobem lze pouze z projevů chování posuzovat jejich míru u jednotlivých účastníků AC? Že se tento problém jenom opustil a nevyřešil dokládá i skutečnost, že se ve prospěch obsahové validity ignoruje koncept konstruktové validity, jenž je například v německy mluvících zemích v souvislosti s využíváním AC úplně neznámý (Krause, Gebert, 2003).

---

## LITERATURA

- Arthur, W., Day, E. A., McNelly, T. L. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56(1), 125–154.
- Arthur, W., Day, E. A., Woehr, D. J. (2008). Mend It, Don't End It: An Alternate View of Assessment Center Construct-Related Validity Evidence. *Industrial & Organizational Psychology*, 1(1), 105–111.
- Arthur, W., Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93(2), 435–442.
- Barrick, M. R., Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Bray, D. W., Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, 80(17), 1–27.
- Bucalo, J. P. (1974). The assessment center: A more specified approach. *Human Resource Management*, 13(3), 2–13.
- Chen, H. (2006). Assessment center: A critical mechanism for assessing HRD effectiveness and accountability. *Advances in Developing Human Resources*, 8(2), 247–264.
- Eurich, T., Krause, D., Cigularov, K., Thornton, G. C. (2009). Assessment Centers: Current Practices in the United States. *Journal of Business & Psychology*, 24(4), 387–407.
- Ferjenčík, J. (2010). Úvod do metodologie psychologického výzkumu. Praha: Portál.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Assessment Center Technology*, 72(3), 493–511.
- Haaland, S., Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology*, 55(1), 137–163.
- Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology*, 64(2), 351–395.
- Howard, A. (2008). Making Assessment Centers Work the Way They Are Supposed To. *Industrial & Organizational Psychology*, 1(1), 98–104.
- Huck, J. R. (1973). Assessment centers: A review of the external and internal validities. *Personnel Psychology*, 26(2), 191–212.
- Van Iddekinge, C. H., Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, 61(4), 871–925.

- International Task Force on Assessment Center Guidelines. (2009). Guidelines and Ethical Considerations for Assessment Center Operations. *International Journal of Selection and Assessment*, 17(3), 244–254.
- Jansen, P. W., Stoop, B. M. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology*, 86(4), 741–753.
- Krause, D. E., Gebert, D. (2003). A Comparison of Assessment Center Practices in Organizations in German-speaking Regions and the United States. *International Journal of Selection & Assessment*, 11(4), 297–312.
- Lance, C. E. (2008). Why Assessment Centers Do Not Work the Way They Are Supposed To. *Industrial & Organizational Psychology*, 1(1), 84–97.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F. Conway, J. M. (2004). Revised Estimates of Dimension and Exercise Variance Components in Assessment Center Postexercise Dimension Ratings. *Journal of Applied Psychology*, 89(2), 377–385.
- Lance, C. E., Woehr, D. J., Meade, A. W. (2007). Case Study: A Monte Carlo Investigation of Assessment Center Construct Validity Models. *Organizational Research Methods*, 10(3), 430–448.
- Lievens, F. (2001a). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, 86(2), 255–264.
- Lievens, F. (2001b). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior*, 22(3), 203–221.
- Lievens, F. (2002). An examination of the accuracy of slogans related to assessment centres. *Personnel Review*, 31(1), 86–102.
- Lievens, F. Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, 86(6), 1202–1222.
- Macan, T. H., Avedon, M. J., Paese, M., Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47(4), 715–738.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Murphy, K. R. (2009). Validity, Validation and Values. *Academy of Management Annals*, 3(1), 421–461.
- Murphy, K. R., Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology*, 50(4), 823–854.
- Petrides, K. V., Weinstein, Y., Chou, J., Furnham, A., Swami, V. (2010). An investigation into assessment centre validity, fairness, and selection drivers. *Australian Journal of Psychology*, 62(4), 227–235.
- Reilly, R. R., Henry, S., Smither, J. W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, 43(1), 71–84.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56(3), 573–605.
- Scroggins, W. A., Thomas, S. L., Morris, J. A. (2008). Psychological Testing in Personnel Selection, Part I: A Century of Psychological Testing. *Public Personnel Management*, 37(1), 99–109.
- Schmidt, F. L., Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and Theoretical Implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46(1), 49–76.
- Taft, R. (1959). Multiple methods of personality assessment. *Psychological Bulletin*, 56(5), 333–352.
- Vaculík, M. (2010). *Assessment centrum: Psychologie ve výběru a rozvoji lidí*. Brno: NC Publishing.

## ABSTRACT ASSESSMENT CENTRE VALIDITY RESEARCH

M. VRANKA

### ABSTRACT

Intuitively appealing concept of an assessment centre and relatively satisfactory success rate of its predictions lead to the long-term neglect of theoretical and methodological problems associated with this selection

tool. Only the last few years have brought an explicit renewal of interest in these fundamental questions and their possible solutions. The aim of the article is to give an overview of the current situation and the results of current research in order to facilitate their application in practice.

**Keywords:** assessment centre, predictive validity, construct validity

## **DIE VALIDITÄT VON ASSESSMENT CENTERN ERFORSCHEN**

M. VRANKA

### **ABSTRAKT**

Die intuitiv ansprechende Konzeption und relativ zufriedenstellende Erfolgsrate der Vorhersagen der Assessment-Center half der langfristigen Vernachlässigung der theoretischen und methodologischen Probleme dieser Methode. Erst in den letzten Jahren wurde das Interesse an diesen grundlegenden Fragen und deren Lösungen erneuert. Ziel des Artikels ist es, einen Überblick über die aktuelle Situation und die Ergebnisse der aktuellen Forschung zu bieten und ihrer Anwendung in der Praxis zu helfen.

**Schlüsselwörter:** Assessment Centrum, Prädiktive Validität, Konstruktvalidität

*Mgr. Bc. Marek Vranka, student doktorského studia Sociální psychologie na FF UK v Praze, e-mail: marek.vranka@ff.cuni.cz.*

Text vychází z nepublikované bakalářské práce: Vranka, M. Posouzení validity využití assessment center. Praha, 2012. Bakalářská práce. VŠE v Praze, Fakulta podnikohospodářská.