

Často bývá reliabilita a s ní související přesnost měření interpretována tak, že pokud bychom žákům zadali stejný test znovu, měli by dosáhnout stejného výsledku. Míra odlišnosti výsledků při opakovaném měření stejným testem je pak mírou chyby měření, a tedy zdrojem nižší reliability testu. Takto odhadnuté reliabilitě se říká **dependabilita**. Při tomto sběru dat se předpokládá, že výkony žáků při obou testováních jsou srovnatelné. Na to se bohužel nelze zcela spolehnout. Při opakovaném testování mohou nekontrolovaně působit dva protichůdné faktory: 1) může klesnout motivace žáků při vyplňování stejného testu znovu, 2) žáci se mohou učit při řešení prvního testu nebo následně po něm při rozhovoru se spolužáky, a tedy nejen rozvíjet měřenou dovednost, ale i zjistit, jak mělo být správně odpovězeno na testové položky. Tyto dva faktory nelze nikdy zcela odstranit. Pokud bychom však od uvedených faktorů odhlédli, dá se odvodit za předpokladů uvedených výše, **že reliabilitu testu lze určit jako korelační koeficient mezi výsledky obou testů**. Při sběru dat je tedy potřeba zajistit, aby výsledky žáků nebyly anonymní, a aby tedy bylo možné po jednotlivých žácích propojit výsledky z prvního i druhého testování.⁹³

Pokud existuje delší časový interval mezi zadáním prvního a druhého testu, mluví se o **stabilitě**, která je též určena jako korelační koeficient mezi výsledky při obou administracích testu. Stabilita ale již nevypovídá jen o reliabilitě testu, ale též o stabilitě dovednosti, kterou daný test měří. A zvláště u testů didaktických vycházíme z toho, že dovednosti, které test měří, mají být školou rozvíjeny, tudíž se stabilita dovedností ani nepředpokládá. Proto se zjišťování stability týká spíše testů psychologických než didaktických.

Další možností odhadu reliability blízkou k dependabilitě je tzv. **ekvivalence**. Ta předpokládá, že existují dvě (minimálně) varianty téhož testu. A obdobně se dá dokázat, že korelační koeficient mezi oběma variantami testu odpovídá teoreticky vymezené reliabilitě. Je tedy potřeba realizovat sběr dat tak, aby vhodně vybranému vzorku žáků, pro které je test určen, byly obě varianty zadány v krátkém časovém odstupu.

Další a nejčastěji používanou metodou odhadu reliability testu je **vnitřní konzistence**. Řeší problém předešlých dvou metod, není potřeba dvojího zadání testu a stačí mít k dispozici jen jeden test a výsledky vhodně vybraného vzorku žáků. Tyto výsledky je potřeba mít po jednotlivých úlohách, nestačí celkový výsledek v testu. Vnitřní konzistence se určuje dvěma postupy:

1. Metodou půlení (tzv. split-half reliabilita)
2. Koeficientem Cronbachovo alfa.

Metoda půlení se používala dříve především z důvodu snazších výpočtů bez pomoci počítače. Je i dobře pochopitelná na základě výše uvedené ekvivalence. Její problém vyplývá z nejednoznačnosti při výpočtu. Na shromážděných datech z testu se výsledky testu rozdělí na dvě poloviny, spočítá se skóre z každé poloviny testu pro každého žáka a reliabilita je odhadnuta jako korelační koeficient mezi výsledky obou částí testu. Tímto ale není odhadnuta reliabilita celého testu, ale pouze jeho poloviny. Jak se dopočítat reliability celého testu, odvodíme ze Spearman-Brownova vzorce (viz níže). Nejednoznačnost spočívá v tom, že rozdělit test na dvě poloviny lze mnoha způsoby. Nejčastěji jsou voleny následující: a) první a druhá část testu, b) liché a sudé položky, c) náhodné přiřazení položek do každé z polovin testu.

⁹³ Zájemce o tento důkaz lze odkázat na publikaci: Zvára, K., & Štěpán, J. (2002). *Pravděpodobnost a matematická statistika*. Praha: Matfyzpress.