

Vývoj nových položek sleduje standardizovaný metodický postup. Primárním krokem je nalezení tzv. stimulu pro tvorbu testových položek. Stimulem je kontext problémové situace, do které je tvorba testových položek zasazena. Poznamenejme, že o vhodnosti generovaných stimulů probíhá diskuse mezi týmy vyvíjejícími testové položky, přičemž v případě shody o vhodnosti daného stimulu je tento rozpracován do podoby navazujících testových položek. V dalším kroku probíhá intenzivní hodnocení kvality vytvořených testových položek. Za tímto účelem jsou využívány následující postupy:

- hodnocení kvality testových položek expertním panelem,
- kognitivní rozhovory s relativně malou skupinou žáků týkající se srozumitelnosti testových položek,
- pilotáž týkající se srozumitelnosti a obtížnosti testových položek.

Uvedené metody jsou aplikovány rovněž v mezinárodním kontextu, a to mimo jiné za účelem poznání vhodnosti testových položek v různých kulturních prostředích. V průběhu celého kontrolního procesu dochází k přeformulování testových položek, ale i k jejich úplnému opuštění. Schválené testové položky jsou přeloženy do jazyků všech účastnících se zemí, přičemž i v tomto stádiu vývoje nových položek dochází k úpravě jejich podoby. Posledním krokem hodnocení kvality nových testových položek šetření PISA je mezinárodní pilotáž na vzorku cca 200 žáků v každé z participujících zemí. Výsledky pilotáže jsou detailně analyzovány prostřednictvím IRT⁸⁴ teorie s cílem identifikovat mimo jiné obtížnost testové položky vzhledem k odpovědím žáků či schopnost testové položky rozlišovat mezi odpověďmi dobrých a špatných žáků.

Při tvorbě výsledných testových sešitů je postupováno tak, že cca 5 testových položek je uspořádáno do položkové jednotky na bázi společného stimulu, tj. popisu kontextové situace ve formě testu, tabulky, obrázku či grafu. Ty jsou následně spojeny do tzv. klastrů jednotlivých testovaných gramotností, přičemž každý klastr obsahuje cca 30 minut testového materiálu. Každý žák dostává pro testování čtyři klastry testových položek, a to dva klastry pro hlavní testovanou gramotnost a po jednom klastru pro zbývající dvě testované gramotnosti. Klastry jsou přitom řazeny do testových sešitů na základě dvou podmínek:

- klastr se objevuje na každém místě pořadí v testovém sešitě (tj. jako 1., 2., 3. i 4. klastr testového sešitu),
- každá dvojice klastrů se objevuje společně jen v jednom testovém sešitě.

Každému žákovi je náhodně přiřazen jeden z takto vytvořených testových sešitů. Poznamenejme, že speciálně vyvíjeny jsou testové sešity pro žáky se speciálními vzdělávacími potřebami respektive testové sešity s nižší náročností. Druhý typ testových sešitů je primárně utvářen pro ty územní celky, jejichž žáci dosahují relativně horších výsledků ve vzdělávání. Přístup na bázi jednodušších testových sešitů umožňuje jednak lépe diferencovat dosažené výsledky a jednak podněcuje vyšší motivaci účastnících se žáků. Konečně uveďme, že šetření PISA pracuje s různými formáty testových položek, od multi-choice výběru se 4 až 5 možnostmi výběru či dichotomickou odpovědí až po otevřené testové položky s krátkou (např. text, slovo) respektive dlouhou odpovědí. Poznamenejme, že poměr těchto tří typů testových položek je třetinový.

⁸⁴ *Item Response Theory*