

- zajištění tisku a distribuce testových materiálů,
- realizace všech nezbytných administrativních úkonů a sběr podnětů od účastníků se škol ke zlepšení NAEP testování,
- zpracování odpovědí žáků do ICT založeného systému sběru dat.

Národní úřad pak má za úkol ve spolupráci se školami vybrat soubor žáků účastníků se NAEP testování. Přiřazení testových materiálů a následné vyhodnocení odpovědí je založeno na identifikačním čísle žáka, a to při respektování principu ochrany osobních údajů (US DOE 2009a).

Realizace NAEP testování probíhá ve spolupráci se školami. Opravu testů zajišťují týmy cca 12 proškolených hodnotitelů pod dohledem supervizora. Oprava testů je realizována s využitím ICT technologií, kdy odpovědi žáků jsou skenovány speciálně připraveným programem, který nabízí hodnotitelům možnosti výběru hodnocení. Tento přístup umožňuje rovněž flexibilní kontrolu správnosti hodnocení ze strany supervizora. Ta probíhá jednak přezkumem 5 % náhodně vybraných položek hodnocených každým hodnotitelem a jednak prostřednictvím analýzy statistických odchylek v hodnocení. Konečně kontrola správnosti hodnocení odpovědí mezi roky vychází ze zařazení určitého počtu odpovědí z minulých let do procesu opravování a porovnání rozdílů v hodnocení (US DOE 2009a).

4. Vyhodnocení testu

Vlastní vyhodnocení NAEP testování probíhá v duchu myšlenky Raschova modelu s využitím DIF¹²² analýzy pro posouzení rozdílů mezi jednotlivými skupinami a IRT teorie¹²³, která utváří teoretické východisko pro převedení výsledků žáků na škály. Škály jsou utvářeny zvláště pro jednotlivé předměty¹²⁴ a neumožňují tak srovnání mezi nimi (US DOE 2010b). NAEP testování není určeno pro hodnocení jednotlivých žáků a soustředí se přímo na hodnocení výsledků vzhledem ke sledovaným skupinám žáků (viz US DOE 2009a):

- např. skupiny žáků podle etnického původu, pohlaví, lokalizace školy,
- např. skupiny žáků podle států s dalším členěním podle jejich charakteristik.

Vyhodnocení dat NAEP testování je založeno na výpočtu jednak průměrného skóre pro každou hodnocenou skupinu žáků a jednak průměrného skóre pro skupiny žáků dosahujících vybraných percentilů výkonnosti – 10, 25, 50, 75 a 90. Vedle toho jsou definovány tři úrovně dosažených výsledků – základní, pokročilý, velmi pokročilý. Tyto úrovně blíže popisují, co žák na dané úrovni výkonnosti umí a na tomto základě jsou stanoveny i meze intervalů hodnotící škály předmětů. Konečně v rámci vyhodnocení NAEP testování jsou rovněž sledovány změny v časovém vývoji. Poznamenejme, že vyhodnocení NAEP testování respektuje základní pravidla inferenční statistiky, kdy rozdíly mezi skupinami či v časovém vývoji jsou zdůrazněny jen tehdy, pokud jsou statisticky významné na pětiprocentní hladině významnosti a dále že relativně vyšší počet škol některých strat ve výběrovém souboru je kompenzován prostřednictvím nižší váhy jejich odpovědí (viz US DOE 2010b).

¹²² *Differential Item Functioning*

¹²³ *Item Response Theory*

¹²⁴ Takto například NAEP testování pro 12. ročník studia v roce 2009 využívalo škálu 0-500 pro čtení a škálu 0-300 pro matematiku (US DOE 2010b).