

Risikofyldte test – erfaringer fra dansk testhistorie

I 1930 påbegyndte psykolog og kommunelærer Henning Meyer sit skolepsykologiske arbejde på Frederiksberg, og dermed lød startskuddet til en ny og skelsættende praksis, der i de følgende år skulle brede sig til hele den danske folkeskole. Et af de afgørende nye elementer i denne praksis var anvendelsen af den danske standardisering af Binet-Simon intelligencetesten, der kom til at fungere som et videnskabeligt forankret værktøj, der viste sig funktionelt i forhold til at sortere børn i henholdsvis »normalklasser« og »værneklasser« baseret på børnenes intelligenskvotient.

Den danske folkeskole havde imidlertid længe været gennemsyret af princippet om, at hver elev skulle prøves og eksamineres i hovedfagene for at kunne rykke op i næste klasse (Christensen, 1934, s. 129). Dette skete ved gennemførelsen af årsprøver og eksamener i såvel mundtlig som skriftlig form. Men når den nye praksis på Frederiksberg alligevel er værd at bide mærke i, er det, fordi den afspejler indmarchen af såkaldte videnskabelige test i den danske skoleverden. Disse test adskilte sig på en lang række områder fra eksamener og prøver, fordi de hævdede at være videnskabelige og objektive. Disse nye test var, som i eksemplet fra Frederiksberg, intelligencetest, men der blev også udgivet standardiserede standpunktstest, som lærere og skoler kunne erhverve til brug for testning af deres elever.

Testene blev i datiden set som et alternativ til eksamener, der blev kritiseret for at være subjektive og uvidenskabelige. Bannerførerne i denne bevægelse for indførelsen af videnskabelige test i folkeskolen var de progressive reformpædagoger, hvilket kan synes paradoksalt med nutidige briller. Reformpædagogerne ønskede at



CHRISTIAN YDESEN

Ph.d., ekstern lektor ved Aalborg Universitet
christianydesen@gmail.com

identificere det enkelte barns potentialer og følgelig indrette undervisningen derefter samt placere barnet på dets »rette« hylde i undervisningssystemet. Til denne brug så reformpædagogerne de nye testtyper som progressive værktøjer til at skabe lige og retfærdige muligheder (Ydesen, 2011, s. 59 ff.).

I denne artikel vil jeg se nærmere på tre centrale eksempler på, hvorledes den danske folkeskole i det 20. århundrede har gjort brug af test, og hvilke konsekvenser disse testpraksisser har haft for børnene. Endelig vil jeg komme med nogle betragtninger over, hvilke konkrete erfaringer der med rimelighed kan udledes af disse historiske praksisser.

Risikofyldte test i dansk skolehistorie

Fra 1930'erne vandt intelligenstagning af problembørn stor udbredelse i den danske skoleverden. Denne testpraksis var forbundet med en betydelig risiko for de testede børn, idet testen et meget langt stykke hen ad vejen blev udslagsgivende for barnets videre skæbne i uddannelsessystemet. De første skolepsykologer, med Henning Meyer i spidsen, angav således retningslinjer for fordelingen af børn i forskellige skoleafdelinger baseret på barnets intelligenskvotient. Normalskolen skulle rumme børn med en intelligenskvotient på mellem 90 og 110. Værneskolen skulle bestå af børn med en intelligenskvotient mellem 70 og 90, mens åndssvageforsorgen var reserveret til børn med en intelligenskvotient under 70 (Ydesen, 2011, s. 55). Intelligenskvotientgrænsen på 70 for børn overført til åndssvageforsorgen blev endda knæsat i et cirkulære fra 1943 vedrørende udmøntningen af skoleloven af 1937¹. Forestillinger om intelligens og intelligenskvotienter spillede således en afgørende rolle i bedømmelsen af børnene og dermed deres videre livsforløb.

Men det var ikke kun skolepsykologernes intelligenstag, der rummede en risiko for det testede barn. På Emdrupborg forsøgsskole, der blev etableret i 1948, søsattes det hidtil mest omfattende test- og evalueringsprogram i Danmark med det overordnede formål at omsætte pædagogiske og psykologiske teorier til praksis. Som led i dette program blev samtlige børn, der var indskrevet ved skolen, testet med Uppsala-skolemodenhedstesten, der blev benyttet til at afgøre graden af barnets skolemodenhed. Testen blev således anvendt til at afgøre, hvorvidt barnet umiddelbart kunne starte i skole, eller om det måtte vente et år. Men Uppsala-testen vandt også udbredelse uden for Emdrupborgs mure. I 1958 blev således ca. 7.000 danske førskolebørn ud af ca. 73.400 testet med Uppsala-testen.

Uppsala-skolemodenhedstesten var en prognostisk gruppetest, som tog sigte på at afdække evner forbundet med intelligens med henblik på at identificere børn, der med stor sandsynlighed ville klare sig dårligt i skolen. Men testen blev også anvendt til at undgå overvægt af problembørn i de nye førsteklaser.

Den praktiske gennemførelse af Uppsala-testen tyder imidlertid på, at der var betydelige fejlkilder til stede i testsituationen. Prøvens resultat afhang af, om det var lykkedes at skabe den rette afslappede atmosfære, af observatørens subjektive bedømmelser, problemer med træthed og manglende koncentration, børnenes kendskab og erfaring med testopgaver og testsituationer samt børnenes forventningsfuldhed og usikkerhed i deres første møde med skolen (Ydesen, 2011, s. 155 ff.). Konsekvensen af en ikke-bestået test blev som regel en udskudt skolestart.

Et tredje eksempel på risikofyldte test er den omfattende testpraksis mellem 1961 og 1971 af grønlandske børn i 12-års-alderen. Testen skulle afklare, hvorvidt det enkelte barn kunne komme på et års skoleophold i Danmark og dermed få adgang til en videregående uddannelse.

Behovet for at udvælge de rigtige grønlandske børn fik skoledirektoratet og Ministeriet for Grønland til at se på test som en brugbar selektionsteknologi, der kunne sikre objektivitet i udvælgelsen. Ud over faglige test i disciplinerne dansk og grønlandsk diktat, stileskrivning samt regning blev fra 1969 også den såkaldte GNVTB (Greenlandic Non-Verbal Test Battery) intelligenstaget anvendt i udvælgelsen af grønlandske børn, om end den dog kun blev anvendt i tvivlstilfælde. Årsagen til inddragelsen af GNVTB-testen var den kritik, som Jesper Florander (1918-1988) fra Danmarks Pædagogiske Institut ved flere lejligheder rejste af de anvendte faglige test samt lærerevalueringerne. Florander såede tvivl om den eksisterende udvælgelsesprocedures evne til at identificere alle egnede grønlandske børn og påpegede et behov for en kulturneutral test. Valget faldt på NVTB-testen, der var udviklet af den danske skolepsykolog og senere professor ved Danmarks Lærerhøjskole Kaj Spelling (1915-1994) under hans arbejde for UNESCO i Malaysia i slutningen af 1950'erne. Testen blev siden standardiseret på 375 grønlandske børn fra 10 byer og 8 bygder i Vestgrønland, hvilket afstedkom tilføjes af G'et i testens betegnelse (Ydesen, 2011, s. 202 ff.).

Inddragelsen af GNVTB-testen i det allerede omfattende testbatteri medførte imidlertid ikke en højere grad af klarhed og objektivitet. På trods af testens sigte som en kulturneutral intelligenstaget indeholdt den fremdeles kulturelle normer og værdier i form af

abstrakt tænkning og evnen til logisk ræsonnement. Den kulturelle skævhed kom blandt andet til udtryk derved, at mange grønlandske børn fandt det vanskeligt at arbejde med abstrakte geometriske figurer, som var en kernekomponent i GNVTB-testen (ibid.).

De historiske erfaringer

De historiske erfaringer med disse tre testpraksisser afslører en række problematiske forhold, som det også i dag er værd at være opmærksom på.

For det første kræver gennemførelsen af en test, at testtageren er rolig, afbalanceret og fattet. Dette har man historisk forsøgt at løse ved at lægge stor vægt på tilvejebringelsen af en rolig og uhøjtidelig atmosfære i selve testsituationen for på den måde at skabe gode arbejdsbetingelser for testtageren. Men det er klart, at dette må være behæftet med en del usikkerhed. For det andet kræver sammenligning af testresultater, at samtlige test er foregået under præcis de samme forhold – et forhold, som ligeledes kan være vanskeligt at tilvejebringe. Og for det tredje er ingen test bedre, end testdesigneren har været i stand til at lave den. Det vil sige, at den logik, der er indlagt i testen, og testens indbyggede kulturelle værdier – det forhold, at noget nødvendigvis må anses for bedre end noget andet for at kunne foretage en evaluering – er af afgørende betydning for kvaliteten. Ligeledes kan paletten af svarmuligheder i en multiple choice-test være afgørende for testens kvalitet. Et relevant spørgsmål til yderligere refleksion er, hvorledes de forkerte svar i en multiple choice-test kan være forkerte efter alle standarder, men samtidig være tillokkende nok til at kunne udfordre det rigtige svar. Endelig betyder også det anvendte sprog og de anvendte begreber noget for testens kvalitet. Årsagen er, at testtageren gerne skulle bibringes den rigtige forståelse af spørgsmålet, og her spiller sproglige koder og begrebslige associationer en central rolle. Alene disse fejkilder peger på, at det ikke er tilrådeligt at bedømme et barn på grundlag af blot en enkelt test.

Et mere fundamentalt problem er, at en standardtest vanskeligt kan indfange individets righoldighed. De prognostiske elementer, som fandtes i alle tre historiske testpraksisser, viste sig at indeholde betydelige usikkerheder. Alle mennesker er naturligvis forskellige, og derfor kan man ikke håbe på, at en standardtest kan give en udtømmende beskrivelse af et menneske eller blot dets evner inden for et bestemt område. I den forbindelse er det væsentligt at være opmærksom på, at testresultaterne i standardiserede test ikke kun

afhænger af testtagerens svar, men i lige så høj grad af normgruppens svar, eftersom testresultaterne bedømmes i forhold til normgruppen.

Endelig viser de historiske erfaringer med test, at deres indførelse er tæt forbundet med tanker om effektivitet, både uddannelsesmæssigt og samfundsøkonomisk. I såvel de historiske eksempler som i den nuværende situation spiller økonomisk effektivitet og opfattelser af den internationale konkurrencesituation en afgørende rolle i forhold til skabelsen af et frodigt klima for test. Det kommer blandt andet til udtryk i erhvervslivets kritik af skolen og af de kundskaber, eleverne forlader skolen med. Denne kritik var også dominerende i mellemkrigstiden, så på denne front er der ikke meget, der har forandret sig. En sådan kritik fremmer planer og ønsker – både i og uden for skolen – om at måle uddannelsessystemets produktion.

På baggrund af disse pointer er det næppe nogen overdrivelse at appellere til en høj grad af ydmyghed i omgangen med testresultater. Appellen understøttes af de historiske erfaringer med anvendelse af testpraksis på både Frederiksberg, Emdrupborg og i Grønland. Det har nemlig vist sig, at test i praksis har en tendens til at underminere og begrænse helhedsvurderinger samt i det hele taget andre evalueringsformer som eksempelvis kvalitative lærerevalueringer. Årsagen er, at det er så besnærende at sammenligne talmæssige størrelser uden hensyntagen til de mange forbehold og fejlkilder, som ligger indbygget i testresultaterne.

Note

- 1 Cirkulære vedrørende gennemførelse af de i folkeskoleloven af 18. maj 1937 §2, stk. 5 givne bestemmelser om indretning af særundervisning for børn, der ikke kan følge den almindelige undervisning, §1-5.

Litteratur

- Christensen, Georg (1934): Examensforhold i Danmark. *Skola och Samhälle* (3), pp. 125-129.
- Ydesen, Christian (2011): *The Rise of High-Stakes Educational Testing in Denmark (1920-1970)*. Frankfurt am Main: Peter Lang Verlag.