**Indirect assessment of practical science skills: Development and application of a taxonomy of written questions about practical work**

**Frances Wilson** OCR, **Stuart Shaw** Cambridge Assessment International Education,  
**Neil Wade** OCR, **Sarah Hughes** OCR, and **Sarah Mattey** Research Division  
(The study was completed when the fourth and fifth authors were based in Cambridge Assessment International Education)

Vurdering af praktiske videnskabelige færdigheder:

Udvikling og anvendelse af en taksonomi af skriftlige spørgsmål om praktisk arbejde

Frances Wilson OCR, Stuart Shaw Cambridge Assessment International Education,

Neil Wade OCR, Sarah Hughes OCR og Sarah Mattey Research Division

(Undersøgelsen blev afsluttet, da den fjerde og femte forfatter var baseret i Cambridge Assessment International Education)

# Baggrund

Praktisk arbejde er centralt i naturvidenskabelig uddannelse og bruges ikke kun til at støtte udviklingen af ​​konceptuel viden, men også for at give studerende mulighed for at udvikle en bred vifte af færdigheder, herunder datahåndtering, eksperimentel design og udstyrsmanipulation (Wilson, Wade,

& Evans, 2016). Praktisk arbejde kan vurderes ved hjælp af mange forskellige former for bedømmelse, herunder kursusprojekter, praktiske eksamener og skriftlige spørgsmål til eksamen. Forskellige former for vurdering kan vurdere forskellige aspekter af dette komplekse domæne. Som sådan er det vigtigt at etablere en klar forståelse af de færdigheder og viden, der vurderes ved hver evalueringsform. I denne artikel fokuserer vi på evalueringen af ​​en metode til vurdering af praktisk videnskab: skriftlige eksamensspørgsmål.

Abrahams og Reiss (2012) definerer praktisk arbejde som "et overordnet udtryk, der refererer til enhver form for videnskabsundervisning og læringsaktivitet, hvor studerende, enten arbejder individuelt eller i små grupper, er involveret i at manipulere og / eller observere virkelige genstande og materialer" (p.1036). Dette er i modsætning til virtuelle materialer, såsom dem, der fås fra en video- eller digital simulering, eller endda fra en tekstbaseret konto. Desuden antyder Lunetta, Hofstein og Clough (2007, s.394), at praktiske aktiviteter kan:

1. opleves i skoleindstillinger, hvor studerende interagerer med materialer for at observere og forstå den naturlige verden;

2. være individuelle eller gruppe- eller store gruppedemonstrationer

3. variere på et kontinuum mellem stærkt struktureret og lærercentreret til åben undersøgelse;

4. vare flere uger, inklusive aktiviteter uden for skolen, eller mindre end 20 minutter;

5. brug et højt instrumenteringsniveau eller overhovedet ikke og

6. inkluderer aktiviteter, hvor studerende indsamler data for at illustrere et princip / forhold (deduktiv), og aktiviteter, hvor studerende indsamler data og forsøger at finde ud af mønstre eller forhold fra dataene (induktiv).

Praktisk arbejde kan derfor omfatte en lang række forskellige aktiviteter, som kan bruges af lærere til at tjene mange forskellige læringsmål. I betragtning af vurderingens indflydelse på, hvad der læres og læres, er det vigtigt at forstå nøjagtigt, hvilke færdigheder der vurderes, og hvordan de forholder sig til de specificerede mål for praktisk arbejde inden for læseplanen. Tidligere arbejde på dette område har artikuleret flere formål til praktisk videnskab, i forskellige granularitetsniveauer og baseret på forskellige fortolkninger af, hvad der anses for at være praktisk arbejde. F.eks. Katalogiserede Kirschner og Meester (1988) primært med videregående uddannelser (HE) 120 forskellige mål for praktisk arbejde, som de syntetiserede til 8 overordnede mål.

Disse mål tildeles imidlertid ikke lige stor betydning af lærere. Derudover kan den relative betydning af målsætningerne for praktisk arbejde på forskellige uddannelsesniveauer og på tværs af de forskellige videnskabelige fag variere (Kerr, 1963).

Internationalt bruges en lang række former for vurdering til at vurdere praktiske færdigheder. Abrahams, Reiss og Sharpe (2013) skelner mellem den direkte vurdering af praktisk arbejde, der inkluderer en praktisk eksamen, hvor studerende observeres, der udfører en praktisk aktivitet, og indirekte vurdering af praktisk arbejde, hvor studerende kan vurderes på produktet fra en praktisk aktivitet (f.eks. en laboratorierapport) eller kan stilles skriftlige spørgsmål i en eksamen. Selvom skriftlige spørgsmål om praktisk arbejde ikke kan vurdere de studerendes evne til at manipulere udstyr, bruges de ofte til at vurdere andre aspekter af praktisk arbejde, herunder viden om eksperimentelle procedurer og teknikker, dataanalyse og præsentation og fortolkning af data med hensyn til videnskabelige begreber.

Et sekundært mål med inkludering af skriftlige spørgsmål om praktisk arbejde er at tilskynde til undervisning og læring af praktisk arbejde, fordi dette vil fungere som en god forberedelse til vurderingen. Dette vil understøtte udviklingen af ​​disse færdigheder (f.eks. Manipulerende evner), som ikke vurderes direkte. Skriftlige spørgsmål kan bruges som en del af en skriftlig eksamen eller omfatte en hel eksamensopgave.

I betragtning af mangfoldigheden af ​​praktiske videnskabelige færdigheder, der kan vurderes, er det vigtigt at forstå, hvilke færdigheder der vurderes, og i hvilken andel af en given vurdering. Lærere og studerende skal vide, hvordan man strukturerer undervisning og læring for at sikre, at de er forberedt på vurderingen. På samme måde skal vurderingsorganisationer (AO'er), der sætter vurderingerne, sikre, at den vurderede færdighedsbalance er passende til det tilsigtede pensum, både inden for en enkelt vurdering og på tværs af en kvalifikations levetid.

Selvom individuelle læseplaner og læseplaner angiver de færdigheder og viden, der skal vurderes i en bestemt kvalifikation (f.eks. Vurderingen af, hvordan observationer registreres, måles og estimeres), er der i øjeblikket ingen rammer for at kategorisere de færdigheder og viden, der vurderes ved hjælp af skriftlige spørgsmål om praktisk arbejde, der er designet til at blive brugt til at sammenligne forskellige kvalifikationer og brugt i forskellige sammenhænge.

Denne artikel beskriver en undersøgelse, hvor udviklingen af ​​en taksonomi først beskrives, derefter dens anvendelse i evaluering af aktuelle videnskabelige kvalifikationer undersøgt. Taxonomien sigter mod at klassificere praktiske færdigheder vurderes ved hjælp af skriftlige spørgsmål om praktisk arbejde.

Taxonomien søger at give en tilgængelig beskrivelse af de færdigheder, viden og forståelse (konstruktioner), der ligger til grund for praktisk videnskab i skriftlige eksamensopgaver. En taksonomi af praktiske videnskabelige færdigheder på spørgsmål om praktisk videnskab har potentialet til at muliggøre evaluering og overvågning af praktiske videnskabsspørgsmål til AO'er, da det muliggør sammenligning af kvalifikationer vurderet over tid, mellem artikler og mellem fag. Cambridge Assessment International Education-evalueringer inkluderer skriftlige praktiske videnskabsspørgsmål (som et alternativ til praktiske eksamener), og en evaluering af, hvordan de vurderede færdigheder har ændret sig over tid, og hvordan de varierer mellem fag, ville give yderligere oplysninger om udførelsen af ​​vurderingerne.

OCR-vurderinger af skriftlig praktisk videnskab blev indført i 2016 som et resultat af reformer i kvalifikationerne. Derfor ville evaluering af, om evalueringerne ligner prøvevurderingsmaterialer (SAM'er), og hvordan fag sammenlignes med hinanden med hensyn til de vurderede kvalifikationer, hjælpe mødet med myndighedskrav til at evaluere og overvåge de nye vurderinger.

Undersøgelsen beskrevet her blev gennemført i to faser:

Fase 1: Udvikling af taksonomien, der først tager sigte på at udvikle en taksonomi, derefter at få feedback fra interessenter for at bestemme dens effektivitet. Det indebar derfor to faser:

1. input fra lærer og interessenter; og

2. Forfining af taksonomien.

Fase 2: Anvendelse af taksonomien havde tre mål:

1. At analysere de færdigheder, der vurderes i forskellige papirer, for forskellige

videnskaber og på tværs af flere år;

2. At bruge data om funktionen af ​​individuelle spørgsmål til at evaluere kvaliteten af ​​spørgsmål, der vurderer særlige færdigheder; og

3. At bestemme i hvilke andre sammenhænge og til hvilke fordele taksonomien kunne anvendes.

Fase 1: Udvikling af taksonomien

Fokus for denne undersøgelse var General Certificate of Secondary Education (GCSE), General Certificate of Education (GCE) Advanced Level (A Level) og Advanced Subsidiary Level (AS Level) videnskabsspørgeskemaer fra en britisk AO og en international GCSE (IGCSE) ) skrevet alternativ til praktiske videnskabelige spørgsmål.

I den første fase af forskningen omfattede de analyserede artikler:

G IGCSE-alternativ til praktisk (ATP) skriftlige forespørgsler fra juni 2014 og 2015 til biologi, kemi og fysik. ATP-papirerne inkluderer eksperimentelle kontekster, der er dækket af den praktiske vurdering af IGCSE-videnskabsvurderinger, men er indirekte vurdering af praktiske færdigheder; og G GCSE og GCE SAM'er for hvert af de tre videnskabelige fag (designet til at afspejle nylige reformer i vurderingen af ​​praktisk videnskab).

Spørgsmålene i disse artikler omtales som indlejrede IAPS (indirekte vurdering af praktiske færdigheder).

## Trin 1: Input fra lærer og interessenter

Lærere blev bedt om at kategorisere praktiske spørgsmål fra videnskabspapirer. Lærer-centreret input blev betragtet som meget vigtigt i betragtning af bekymring over virkningen af ​​at vurdere videnskabets praktiske arbejde gennem

skriftlige spørgsmål om praktisk arbejde i klasseværelset. En gruppe på ni videnskabslærere blev ansat (tre biologi, tre kemi og tre fysik). Lærere havde forskellige erfaringer med at undervise i GCE AS-niveau, GCSE og IGCSE.

Fase 1 bestod af to sessioner: en emnegruppesession og en tværfaglig gruppesession.

Emne-gruppe session

Lærere blev forpligtet til at arbejde i grupper af trekanter i fagspecialismer (f.eks. Tre biologi-specialister, der arbejder sammen). Hver gruppe blev forsynet med pakker med spørgsmål, der var tilpasset deres specialitet, med spørgsmål om praktisk arbejde fremhævet. Lærerne blev bedt om at overveje, hvordan de ville kategorisere de fremhævede spørgsmål, med hensyn til den type viden eller evner, der blev vurderet, og at kommentere eventuelle aspekter, der vedrører progression fra GCSE og IGCSE til AS-niveau. Efter at have besluttet en liste over kategorier, blev hver gruppe derefter forpligtet til at generere en kort beskrivelse af kategorien, således at en anden naturfaglærer kunne bruge deres kategorier til at identificere spørgsmål, der faldt i hver kategori. Lærere blev også bedt om at anføre eventuelle spørgsmål i deres pakker med papirer, der faldt i hver kategori. Endelig blev lærere opfordret til at logge eventuelle problemer eller udfordringer, der opstod, mens de arbejdede, især i relation til spørgsmål, der var vanskelige at kategorisere.

## Trin 2: Forfining af taksonomien

Fase 2 bestod af at få feedback til den destillerede taksonomi (konstrueret i fase 1) fra interessentgrupper (inklusive repræsentanter fra HE, fagforeninger og lærere). Disse grupper identificerede behovet for adskillelse af tegnefærdigheder i to separate kategorier. For det første den nøjagtige repræsentation (og mærkning) af observerede objekter, såsom krævet i biologi eller geologi, og for det andet den mere abstrakte diagrammatiske repræsentation af objekter ved hjælp af definerede symboler, såsom elektriske kredsløb, molekylære strukturer eller laboratorieudstyr.

Den destillerede taksonomi udviklet i trin 1 blev også undersøgt af vurderingsspecialister inden for hvert emne for at kategorisere vurderingsemner til en efterfølgende kvalitativ analyse. I udførelsen af denne aktivitet

eventuelle vanskeligheder, der blev identificeret ved at tildele en kategori til et spørgsmål eller en vare, blev katalogiseret.

Generel feedback fra interessenter samt kommentarer om taksonomien, der blev indsamlet under et videnskabsforum arrangeret af en britisk AO, bidrog til forfining af udsagn og forklaringer til at producere den endelige taksonomi. Evalueringsspecialister / forumdeltagere bidrog til en række justeringer af taksonomien. Dette var især vigtigt, da taksonomien var anderledes end tidligere kategoriseringer af praktisk arbejde, der er afbildet i litteraturen: Det adskiller sig fra sådanne overskrifter som planlægning, iagttagelse, analyse og evaluering - etiketter, der beskriver arten af ​​et spørgsmål snarere end de færdigheder, som kan bruges i dens færdiggørelse. Denne gennemgang af taksonomien,

med en række interessenter, lærere, universitetslæsere og repræsentanter for det videnskabelige samfund (via Science Forum), gav også feedback, der førte til tilføjelsen af ​​den abstrakte diagrammatiske repræsentation af elektriske kredsløb og stiliserede to-dimensionelle diagrammer af kemisk apparatur som et andet separat tegnefærdighed.

Den endelige taksonomi er vist som tabel 1.

Tre centrale spørgsmål blev rejst af aktiviteterne til dette punkt:

1. Opdelingen af ​​aktiviteter relateret til data i fire forskellige kategorier;

2. Forskellen mellem at tegne en nøjagtig repræsentation og brugen af ​​symboler; og

3. Taxonomiens art.

Lærere havde klart identificeret en række forskellige færdigheder, der var relevante for

brugen af ​​data. Derfor kunne vi identificere spørgsmål, der var direkte knyttet til disse færdigheder. Eksempler på kategorierne kan demonstreres i følgende spørgsmål:

Eksempel 1: Optagelse af data / Datahåndtering

Dette element med 4 markeringer (vist i figur 1) kræver (i), at kandidaten indfanger de relevante data fra grafen, og (ii) at de håndterer dataene for at beregne gradienten og dermed bestemme metodens Young-modul.

## Tabel 1

Diagrammer (apparater og kredsløb)

Repræsentation af evnen til at repræsentere kredsløb udstyr eller kredsløb eller apparater i accepterede former. ved hjælp af accepterede symboler.

Tegning (biologi) og mærkning

Biologisk tegning, nøjagtig repræsentation.

Evnen til nøjagtigt at repræsentere observerede genstande og til at mærke dem korrekt.

Foretagelse af målinger  
Læsning af en skala fra et diagram (fremherskende i alternativ til praktiske artikler).

En praktisk test af evnen til at bruge en skala til at læse. I mere komplekse scenarier kan der gengives et tidsforløbsfotografi eller et skala-diagram, der kræver, at de studerende foretager målinger.

Genkaldelse. Anvendelse af viden.

Et svar på et spørgsmål, der vedrører en praktisk aktivitet defineret i specifikationen, trækker på tilbagekaldelse af teori eller udfører aktiviteten.

Optagelse Observering og læsning af datadata, fortolkning af data

fra en tabel eller en graf, der optager data.

Dette adskiller sig fra at foretage målinger. I dette tilfælde er dataene hentet fra en tabel eller graf. I modsætning til at tegne en graf,

i dette tilfælde henter den studerende data fra grafen. Færdigheden anses også for at involvere evnen til at registrere dataene korrekt.

Dataproblemløsning. analyse

Efter at have identificeret de relevante data fra en række kilder eller forskellige beregninger, der forbinder informationerne for at tillade løsning af et problem.

Håndtering af datakalkulation (f.eks. Beregning gradient).

Brug af data til at udføre en beregning ved hjælp af en formel, der kan kræve omarrangering eller sammenkobling af formler.

Datatolkning / identificering af tendenser

Plotter og fortolker grafer.

Overførsel af data fra tabel til grafisk format. At kunne identificere tendenser fra grafen (fortælle historien om grafen).

Datakvalitet

Evaluering af data og konklusioner.

Sammenligning af resultatet af en aktivitet med det forventede eller accepterede resultat. Brug af numeriske processer til at kommentere datakvaliteten med mulig henvisning til usikkerheden i processen eller målingerne.

Eksperimentel metode,

planlægning og designprocedure inklusive identifikation af variabler.

En forståelse af de involverede processer, der muliggør identifikation af variabler og evnen til at foreslå et eksperiment for at demonstrere den påkrævede hypotese eller resultat.

Forudsigelse af resultater

Forståelse af processer.

Brug af viden og forståelse af en proces til at forudse det sandsynlige resultat af en given rækkefølge af begivenheder. Ofte undersøgt ved at give et scenarie for en praktisk aktivitet, specificere en ændring i omstændighederne og bede om identifikation af eventuelle ændringer i resultatet.

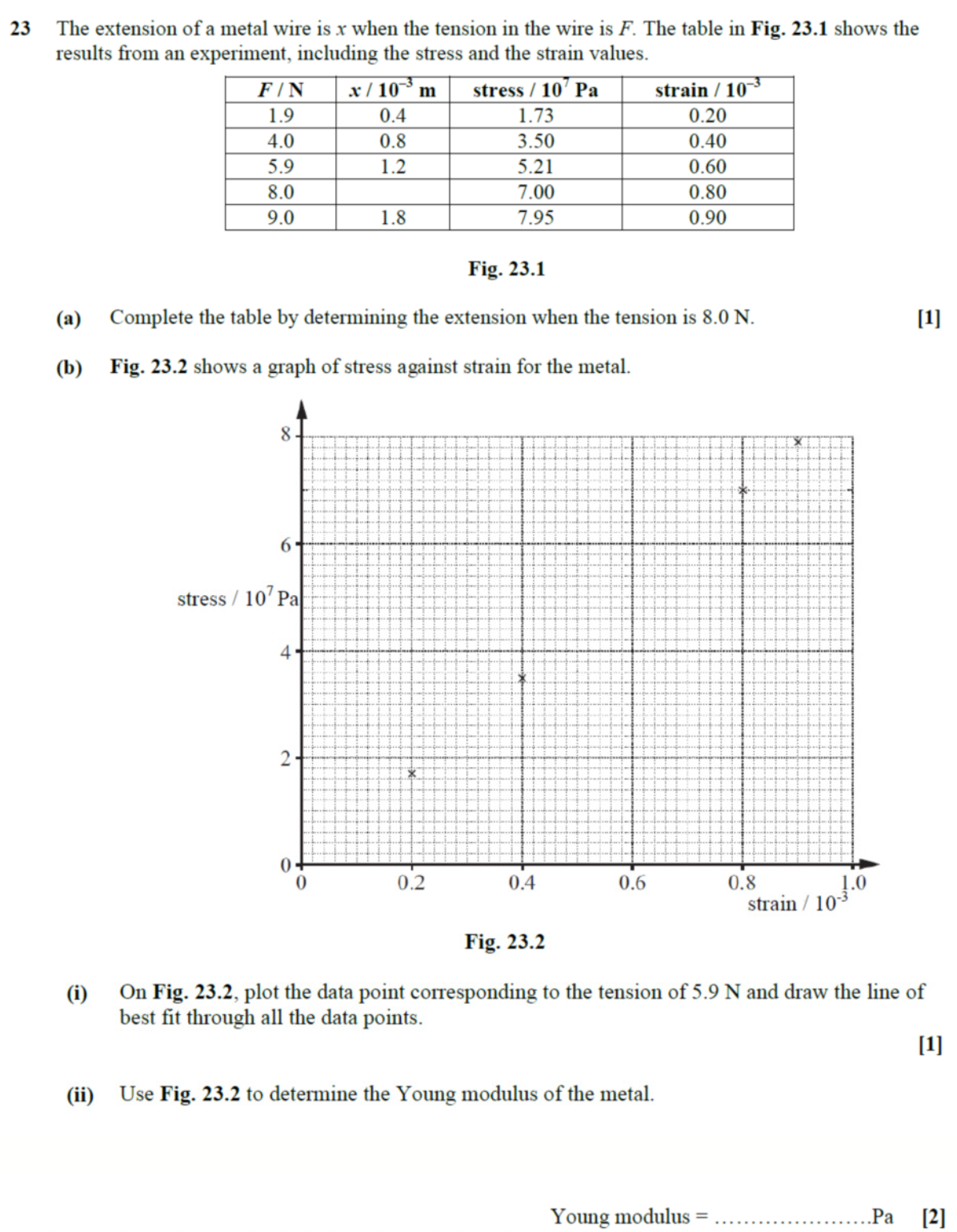
Anvendelse af apparater og teknikker

Anvendelse af viden om praktiske færdigheder.

Typiske spørgsmål kan være evaluering af brugen af specifikt udstyr eller forslag til mulige forbedringer. Dette har antydet forståelse snarere end lige tilbagekaldelse.

## Eksempel 1: Optagelse af data / Datahåndtering

Dette element med 4 markeringer (vist i figur 1) kræver (i), at kandidaten indfanger de relevante data fra grafen, og (ii) at de håndterer dataene for at beregne gradienten og dermed bestemme metodens Young-modul.



## Eksempel 2: Datatolkning / identificering af tendenser

Spørgsmålet i det næste eksempel (vist som figur 2) er typisk for dem, der ønsker at vurdere forståelsen af en praktisk aktivitet ved at kræve, at kandidaten identificerer tendensen for en graf, der matcher det definerede scenarie. Den anden hovedspørgsmålstype i datatolkningskategorien er ved overførsel af tabellata til grafisk form, som inkorporeret i afsnit (b) (i) i eksempel 1 (vist i figur 1).

Et billede, der indeholder tekst, kort

Automatisk genereret beskrivelse

## Eksempel 3: Datakvalitet

Dette er et åbent spørgsmål (vist i figur 3) med i alt

6 mærker til rådighed. Det indeholder en række varer med lavere ordre fra

taksonomien, som derefter bidrager til den højere orden. Spørgsmålet indeholder betydelige data og bruger tidligere spørgsmål, der vedrører de data og grafer, der leveres for derefter at bidrage til den øgede efterspørgsel i sammenligningen mellem de to mulige metoder.

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

## Eksempel 4: Tegning (biologi)

Gennemgangen af en tidlig iteration af taksonomien identificerede evnerne i forbindelse med tegning, evaluering og forklaring som manglende genstande. Quillin og Thomas (2015) diskuterer, hvordan tegninger varierer i det omfang, de er repræsentative eller abstrakte, både inden for biologi og videnskab. Det følgende spørgsmål (vist i figur 4) skildrer en biologisk tegning som en nøjagtig repræsentation af et objekt

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

## Fase 2: Anvendelse af taksonomien

I den anden fase af forskningen blev der anvendt et separat sæt papirer (dem, der blev taget af studerende i 2016). Dette sæt papirer var den første vurdering af de nyudviklede internationale læseplaner, skønt der ikke var nogen ændringer i vurderingen af ​​praktiske færdigheder. Tilføjelsen af ​​spørgeskemaerne i 2016 muliggjorde således analyse af de nuværende britiske og internationale vurderingsmodeller.

Ansøgningsfasen forsøgte at undersøge, om taksonomien kunne anvendes på emner på skriftlig praktisk videnskab inden for tidligere spørgeskemaer og SAM'er. Analysen ved hjælp af taxonomien havde til formål at demonstrere, hvordan taxonomien kunne bruges til at evaluere og overvåge videnskabelige praktiske spørgsmål ved at tage to forskningsspørgsmål op:

1. Vurderes færdighederne i taksonomien?

Dette spørgsmål blev behandlet ved at sammenligne dækningen af ​​færdigheder vurderet i IGCSE, GCSE og SAM-emner på spørgeskemaer på tværs af år og emner. Denne anvendelse af taksonomien giver AO'er mulighed for at overvåge og evaluere, om vurderinger er ensartede i vurderingen af ​​praktiske videnskabelige færdigheder.

2. Hvordan fungerer spørgsmål, der vedrører bestemte færdigheder?

Dette spørgsmål blev behandlet ved at sammenligne, om særlige færdigheder fra taksonomien er forbundet med særlige egenskaber ved varens ydeevne. Denne anvendelse af taksonomien giver AO'er mulighed for at evaluere, om bestemte færdigheder har forskellige egenskaber med hensyn til vanskeligheder og egnethed af emnet inden for resten af ​​spørgeskemaet. For eksempel, hvis visse færdigheder har højere vanskelighedsværdier end andre, kan AO'er muligvis bruge denne information til at uddanne spørgsmålssættere til at anerkende, hvor vanskelige særlige færdigheder er, eller til at bestemme, hvordan færdigheder skal vurderes i fremtiden. Disse analyser giver også bevis for, at de færdigheder, der vurderes, er passende til vurderingen, idet de ikke fører til høje undladelsesniveauer eller har en større vanskelighed end forventet.

## Forskningsspørgsmål 1: Bedømmes færdighederne i taksonomien?

For at tackle dette spørgsmål gennemgik fagspecialister fra både OCR og Cambridge Assessment International Education de spørgsmål, der er anført nedenfor. De gennemgik først hvert emne og vurderede, om det vurderede en praktisk videnskabsfærdighed, og i bekræftende fald identificerede de derefter hvilke færdigheder der blev vurderet ved hjælp af taksonomien. Flere færdigheder kunne tildeles for hvert emne, og den tildelte første færdighed blev brugt som den primære færdighed, som emnet vurderede.

* IGCSE Alternative to Practical (ATP) skrevne forespørgsler fra juni 2014 - 2016 til biologi og kemi og juni 2013, 2015 og 2016 for fysik. Disse benævnes ATP-spørgsmål.
* GCE AS Skriftlige forespørgselsdokumenter for juni 2016 for hvert af de tre naturfaglige fag (både specifikationer A og B for biologi og kemi og specifikation B for fysik). Derudover blev SAM'er til biologi og kemi også anvendt. Disse kaldes indlejrede IAPS-spørgsmål.

Da fagspecialisterne i denne fase anvendte taksonomien på emner, fandt de, at de vurderede praktiske videnskabelige færdigheder i tidligere spørgsmål. papirer og analyser blev udført på de identificerede poster. For det første blev antallet af hver praktisk færdighed, der blev vurderet fra alle spørgsmål, talt, og der blev foretaget sammenligninger på tværs af år, AO'er og fag. Ud over de planlagte sammenligninger blev yderligere sammenligninger undersøgt med hensyn til sammenligning af varetariffer for varer, der vurderer forskellige kvalifikationer, som en form for yderligere oplysninger, der kunne bruges til at evaluere evalueringens resultater. For det andet blev data på vareniveau indsamlet for alle undersøgte spørgeskemaer og brugt sammen med de færdigheder, der blev tildelt til hver artikel for at sammenligne ydelsesmål for genstande i henhold til de forskellige færdigheder, de vurderede.

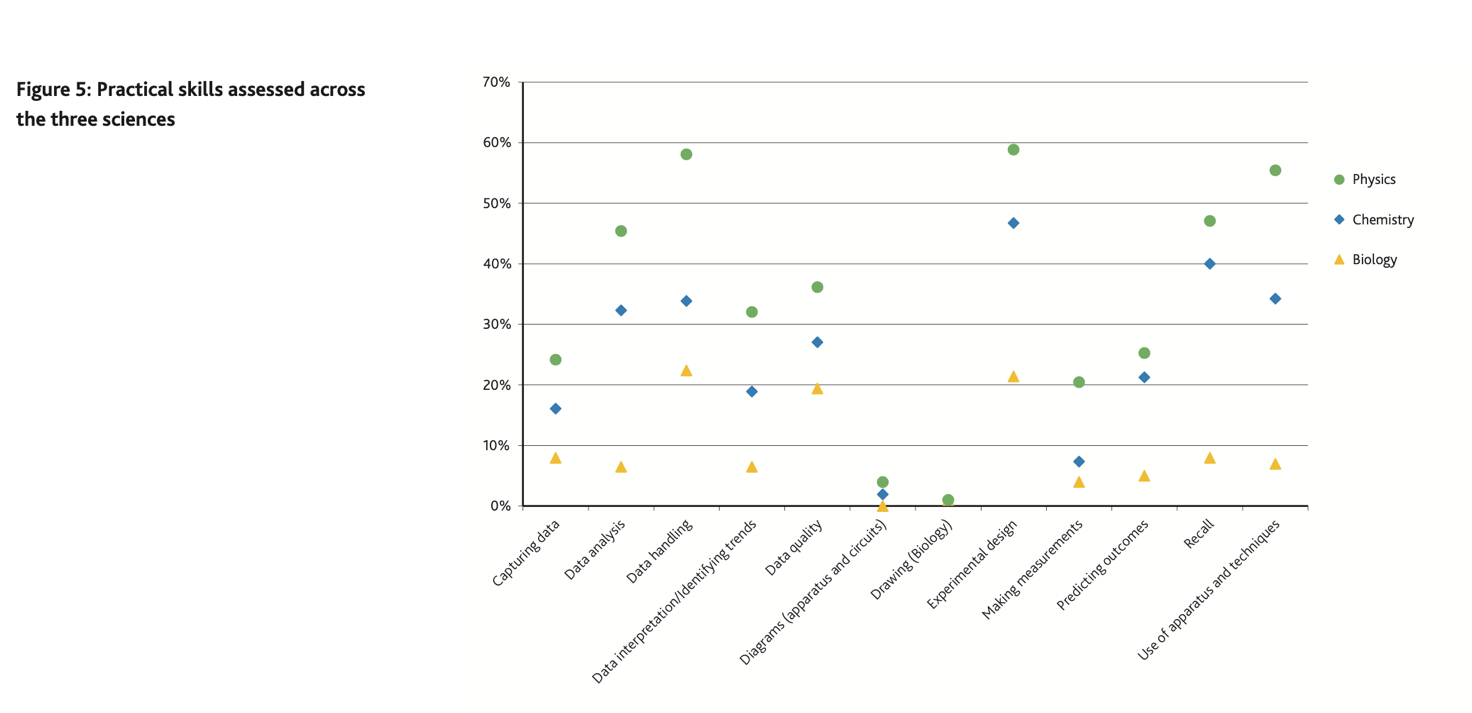
Alle færdighederne inden for taksonomien blev vurderet til at blive vurderet til en vis grad på hvert analyseret papir, skønt de i forskellig grad (se figur 5). De mest almindelige vurderede færdigheder var Eksperimentel design, Datahåndtering, Brug af apparater og teknikker og Genkald. Recalls popularitet påvirkes sandsynligvis af det faktum, at halvdelen

af de spørgsmål, der vurderer denne evne, vurderer også andre færdigheder. Evalueringen af ​​de meget specifikke færdigheder i diagrammer (apparater og kredsløb) og tegning (biologi) var sjælden.

I figur 5 viser vi en sammenligning af de færdigheder, der er vurderet af de tre videnskabelige fag Fysik, kemi og biologi.

Den andel, i hvilken kompetencer vurderes i SAM’erne og live 2016-papirerne sammenlignes i figur 6. Dette muliggør sammenligning af 2016-livevurderingen med designkriterier, der er eksemplificeret i SAM’erne.

Andelene er meget ens, med nogle afvigelser set i dataanalyse, indfangning af data (færre i SAM'er) og datakvalitet (højere procentdel i SAM'er).



Et billede, der indeholder objekt

Automatisk genereret beskrivelse

## Forskningsspørgsmål 2: Hvordan fungerer spørgsmål, der vedrører bestemte færdigheder?

For at undersøge, hvordan spørgsmål, der vedrører bestemte færdigheder, udføres, analyserede vi varens ydelse ved hjælp af de tre mål, der er beskrevet iTabel 2 for hver færdighed og sammenlignede dem med hinanden. Da der ofte blev tildelt flere færdigheder til et emne, brugte vi den tildelte primære færdighed hver vare. Vi brugte facilitet, udeladelse og R\_rest-værdier beregnet for alle elementer, der blev tildelt en praktisk færdighed fra både ATP og indlejrede IAPS-papirer. SAM'er blev ikke brugt, da data på vareniveau ikke er tilgængelige for disse papirer.

For at sammenligne hvor godt elementer, der målretter hver udført færdighed, blev gennemsnittet og standardafvigelsen for facilitet, udeladelse og R\_rest-værdier for alle poster, der blev tildelt til hver færdighed, beregnet. Figur 7 viser den gennemsnitlige ydeevne for hver færdighed med hensyn til facilitet (figur 7a), udeladelsesfrekvens (figur 7b) og R\_rest (figur 7c), hvilket viser variation i facilitet og udeladelsesværdier afhængigt af den vurderede evne.

Facilitets- og udeladelsesværdier antyder to brede grupper af færdigheder, der varierer afhængigt af deres sværhedsgrad. Her har færdighederne Udførelse af målinger, tegning (biologi), indsamling af data, datahåndtering og anvendelse af apparater og teknikker de højeste gennemsnitlige facilitetsværdier og de laveste gennemsnitlige udeladelsesværdier for færdighederne, hvilket indikerer, at ting, der er tildelt som vurdering af disse færdigheder, er lettere sammenlignet med andre praktiske færdigheder.

I modsætning hertil har de færdigheder, dataanalyse, datatolkning / identifikation af tendenser, datakvalitet, eksperimentel design og forudsigelsesresultater de laveste gennemsnitlige facilitetsværdier, og de højeste gennemsnitlige udeladelsesværdier for færdighederne, hvilket indikerer, at emner, der er tildelt som vurdering af disse færdigheder, er vanskeligere sammenlignet med andre praktiske færdigheder.

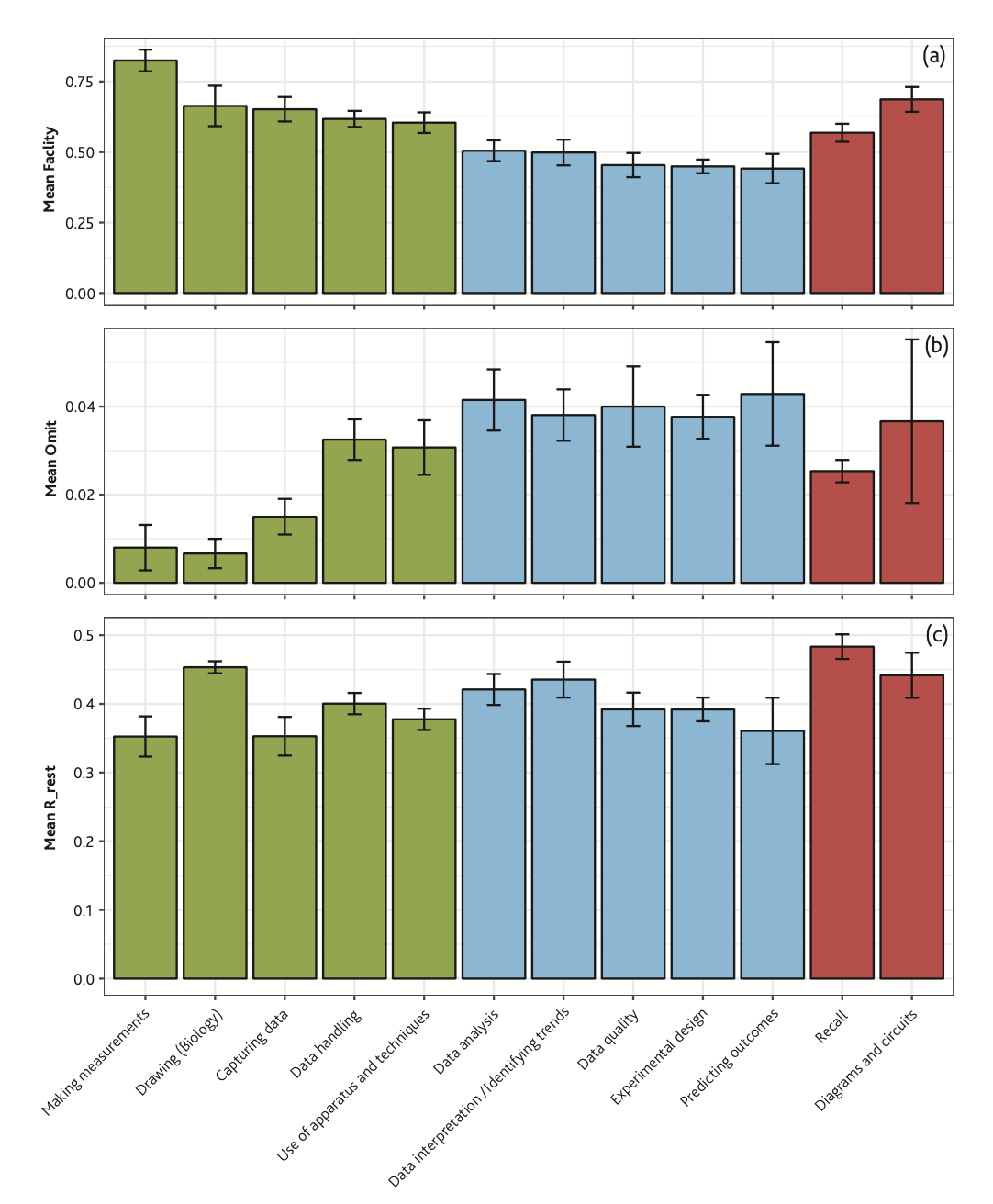
Endvidere passer de elementer, der er tildelt til vurdering af tilbagekaldelse og diagrammer og kredsløb, ikke godt med nogen af ​​grupperne baseret på deres gennemsnitlige facilitet og udeladelsesværdier, og tildeles således ingen af ​​gruppen.

For det første kan dette for færdighedsdiagrammer og kredsløb skyldes den høje variation og et meget lille antal poster, der er tildelt til vurdering af disse færdigheder.

Dernæst kan det, for færdigheden Husk, skyldes den brede vifte af spørgsmål, hvor denne type færdigheder sandsynligvis vil blive vurderet. Endelig er der lidt variation i de gennemsnitlige R\_rest-værdier for hver færdighed, hvilket indikerer, at alle færdigheder klarer sig til et lignende niveau med hensyn til at skelne mellem kandidater.

Mens produktets præstationsmålinger varierede mellem fag (biologi, kemi, fysik), var der intet konsistent mønster, hvorved visse færdigheder i visse fag varierer i sammenligning med andre fag

eller færdigheder.



Figur 7: Målinger af, hvor godt elementer, der vurderer forskellige færdigheder, fungerer efter emne, ved hjælp af både Cambridge International og OCR-data

(Bemærk: Middelfacilitetsværdier (a), gennemsnitlige udeladelsesværdier (b) og gennemsnitlige R\_restværdier (c)) med standardfejlbjælker. Grupperne adskilles ved farve, hvor gruppen med høj facilitet og lav udeladelse er i grønt, gruppe med lav facilitet og høj udeladelse er i blåt, og de færdigheder, der ikke passer i nogen af grupperne, er i rødt).

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

## Diskussion

Det her beskrevne arbejde udgør et forsøg på at udvikle og anvende en taksonomi af skriftlige spørgsmål om praktisk arbejde. Lokalisering af praktiske videnskabelige færdigheder inden for rammerne af en eksplicit ramme giver en mere systematisk og overordnet sammenhængende tilgang til klassificering og konceptualisering af sådanne færdigheder i skriftlige eksamensopgaver.

Integriteten af ​​en praktisk videnskabelig prøve - uanset om den forsøger at vurdere praktiske færdigheder direkte, eller om sådanne færdigheder indirekte udledes (som de måtte være i en skriftlig eksamen eller gennem en anden sekundær form for vurdering) - afhænger i vid udstrækning ved en forståelig forståelse og artikulering af den / de underliggende konstruktion (er), som den søger at karakterisere. Hvis disse konstruktioner ikke er veldefinerede, vil det være vanskeligt at understøtte de påstande, en testudvikler muligvis ønsker at fremsætte om anvendeligheden af ​​testene, herunder påstande om, at videnskabstestene ikke lider af faktorer som konstruer underrepræsentation og konstruer irrelevance ( American Educational Research Association, American Psychological Association og National Council on Measuring in Education, 2014, s.63).

Der skal sættes en nyttig (og nødvendig) sondring mellem hvad er let at vurdere, og hvad der er vigtigt at vurdere. Det kan argumenteres,for eksempel at datafærdigheder (indhentning af data, dataanalyse, datahåndtering og datakvalitet) let vurderes, mens de også betragtes som afgørende for området praktisk videnskabeligt arbejde. Andre færdigheder, der er underlagt klassificeringen Begrebsforståelse, betragtes ofte som en vigtig grund til at udføre praktisk arbejde, skønt interessant er litteraturen om effektiviteten af ​​det praktiske videnskabelige arbejde i at øge udviklingen af ​​begrebsmæssig forståelse noget blandet (Hewson & Hewson, 1983; Hofstein & Lunetta, 1982; Lazarowitz & Tamir, 1994; Mulopo & Fowler, 1987).

Abrahams og Reiss (2015) siger, at indirekte vurdering af praktisk videnskab er mere passende til at bestemme den studerendes forståelse af en færdighed eller fremskridt, mens direkte vurdering er mere passende til at bestemme den studerendes kompetence. Betydningen er, at der er en potentiel fare for, at forståelse af færdigheder, der er lette at vurdere, hyppigt vurderes på bekostning af forståelsen af ​​andre færdigheder og på bekostning af kompetence. Hvis dette er tilfældet, afhængigt af formålet med vurderingen, kan det være nødvendigt at bruge en alternativ vurderingsform i kombination med den skriftlige eksamen.

## Hvordan kan taksonomien bruges, og af hvem?

Vi mener, at taksonomien kan bruges på forskellige måder og af en række uddannelsesudøvere (som lærere, AO'er, læseplanudviklere og testudviklere) til:

* tilvejebringe en struktur til klassificering af etablerede, forudbestemte kategorier af indirekte praktiske videnskabelige færdigheder, der kan bruges af AO'er og testudviklere til at overveje deres intentioner med hensyn til de vurderingspåstander, de ønsker at fremsætte; om en evaluering (fra testudviklerens side) af, hvor effektivt vurderingskravene er blevet opfyldt;
* give testudviklere mulighed for at konstruere spørgsmål, der fremkalder en række og balance mellem passende og effektive praktiske videnskabelige færdigheder;
* sikre et passende niveau af forudsigelighed for dem, der skriver spørgsmål (både på spørgsmål og papirniveau);
* overvåge spørgeskemaer over tid;
* genererer fra et lovgivningsmæssigt perspektiv SAM'er, der er repræsentativ for fremtidige live papirer;
* give testudviklere mulighed for at overveje deres intentioner med hensyn til sammenlignelighed både inden for kvalifikationer (f.eks. På tværs af videnskabelige fag) og mellem kvalifikationer (f.eks. GCSE og dets internationale modpart);
* muliggør, at effektiviteten af formative opgaver bestemmes (og evalueres) i forhold til lærerens formål og mål; og tilbyder testdesign og udviklingsudøvere et middel til at evaluere vurderinger fra forskellige (konkurrent) AO'er.

Gennemsnit R\_rest Gennemsnit udelade gennemsnitlig faclity

* Foretagelse af målinger Tegning (biologi)
* Optagelse af data
* Datahåndtering
* Anvendelse af apparater og teknikker
* Dataanalyse
* Datatolkning / identificering af tendenser
* Datakvalitet
* Eksperimentelt design Forudsigelse af resultater
* Husk Diagrammer og kredsløb

## Reflekterer over, hvordan taksonomien er blevet brugt

Siden udviklingen af ​​taksonomien er det blevet brugt i OCR til at sammenligne dækningen af ​​vurdering af færdigheder på tværs af de tre videnskaber rutinemæssigt. Dette har vist sig nyttigt, idet det har givet oplysninger til vurderingsspecialister, som kan hjælpe med at overveje konsistensen af ​​dækning af færdigheder på tværs af emnerne. Taxonomien er også blevet brugt til at sammenligne kvalifikationsdækning i SAM'erne og livevurderinger. For eksempel er SAM'erne for nylig blevet sammenlignet med A-niveauet i 2017. Som SAMS eksemplificerer for skoler typer spørgsmål og dækning i den levende vurdering,

det er meget vigtigt, at live-vurderingen afspejler den dækning, der er eksemplificeret i SAM'erne. I denne sammenhæng har taksonomien vist sig at være uvurderlig for OCR. Mere generelt kan AO'er bruge denne information til at støtte et godt match mellem SAM'erne og levende materialer og derved støtte skoler i deres forståelse af forventningen til vurderingerne.

Taxonomien har bidraget til at etablere en klar og velartikeliseret holdning til den / de underliggende konstruktion (er) af skriftlige praktiske videnskabelige vurderinger. At have en klar forståelse af, hvordan praktiske videnskabskonstruktioner er konceptualiseret, vil tjene til at operationelisere disse konstruktioner til vurderingsformål i fremtiden. Som sådan vil taksonomien understøtte efterfølgende omudvikling og / eller revision af videnskabskvalifikationer og give et forsvarligt grundlag for de foreslåede ændringer på konstruktion og andre grunde såsom praktisk, indflydelse, gyldighed og pålidelighed.

Konklusion

Taxonomien giver rammer for overvejelse af, hvilke praktiske færdigheder der vurderes, og hvor ofte de vurderes. Imidlertid behandler det ikke det mere grundlæggende spørgsmål om, hvilke praktiske færdigheder vi skal vurdere i en skriftlig opgave, og hvilken relativ hyppighed af vurderingen hver evne skal have. I denne undersøgelse har vi sammenlignet emner, der blev brugt til at vurdere to forskellige niveauer (GCSE og GCE) og på tværs af biologi, kemi og fysik. Bør GCSE / IGCSE-studerende vurderes på samme færdigheder (og i samme proportioner) som GCE-studerende? GCSE / IGCSE tjener uden tvivl et bredere formål (udvikling og udvikling af videnskabelig færdighed) end GCE,

som er fokuseret på forberedelse til HE. Skal biologi, kemi,

og fysik følger lignende mønstre, eller indebærer tværfaglige forskelle, at forskellige proportioner er passende? Det er uden for denne artikels rækkevidde at besvare disse spørgsmål. Ved at overveje taksonomien i forhold til formålet med en videnskabelig praktisk vurdering, mener vi, at vi har leveret et værktøj til at ramme debatten op.

Tak

Vi er gældsatte over for et antal mennesker, der har bidraget til denne undersøgelse. Vi takker især Diana Marosi, tidligere Cambridge Assessment International Education, for hendes analytiske input. Vi vil også gerne takke de vurderingslærere og lærere, der er involveret i studiet, for deres input gennem hele arbejdet.

References

Abrahams, I., & Reiss, M. J. (2015). The assessment of practical skills. *School Science Review, 96*(357), 40–44.

Abrahams, I., & Reiss, M. J. (2012). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching, 49*(8), 1035–1055.

Abrahams, I., Reiss, M. J., & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education, 49*(2), 209–251.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Hewson, M., & Hewson, P. (1983). Effect of instruction using student prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching, 20*(8), 731–743.

Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research, 52*(2), 201–218.

Kerr, J. F. (1963). *Practical work in school science*. Leicester: Leicester University Press.

Kirschner, P. A., & Meester, M. A. M. (1988). The laboratory in higher science education: Problems, premises and objectives. *Higher Education, 17*(1), 81–98.

Lazarowitz, R., & Tamir, P. (1994). Research on using laboratory instruction in science. In Gabel, D. L. (Ed.), *Handbook of research on science teaching and learning*. (94–128). New York: Macmillan.

Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and teaching in the school science laboratory: An analysis of research, theory and practice. In S. K. Abel & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 393–441). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mulopo, M. M., & Fowler, H. S. (1987). Effects of traditional and discovery instructional approaches on learning outcomes for learners of different intellectual development: A study of chemistry students in Zambia. *Journal of Research in Science Teaching, 24*(3), 217–227.

Quillin, K., & Thomas, S. (2015). Drawing-to-Learn: A Framework for Using Drawings to Promote Model-Based Reasoning in Biology. *CBE Life Science Education. 14*(1):es2.

Wilson, F., Wade, N., & Evans, S. (2016). Impact of changes to practical assessment at GCSE and A-level: the start of a longitudinal study by OCR. *School Science Review, 98*(362), 119–128.