

Een beoordelingsinstrument voor practicanten in de werkplaatstechniek

Citation for published version (APA):

Crombag, H., Meuwese, W. A. T., & Bartholomeus, H. (1965). *Een beoordelingsinstrument voor practicanten in de werkplaatstechniek*. (TH Eindhoven. Onderafd. Wijsbegeerte en Maatschappijwetenschappen. Onderwijsresearch : rapport; Vol. 8). Technische Hogeschool Eindhoven.

Document status and date:

Gepubliceerd: 01/01/1965

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.



Een beoordelingsinstrument voor
praktikanten in de werkplaatstechniek

H. Crombag
W. Meuwese
M. Bartholomeus

Rapport nr. 8
juni 1965

H. Crombag, W. Meuwese, en H. Bartholomeus

Tijdens het eerste semester van het eerste studiejaar volgen de studenten in de werktuigbouwkunde het praktikum werkplaatstechniek. Het doel van dit praktikum is de studenten vertrouwd te maken met de vormgevingsmogelijkheden en de constructie van de verschillende machines in de werkplaats. De studenten worden per middag in groepjes van zes verdeeld over de instructeurs. Het duurt 10 middagen. De instructeurs van het praktikum zijn geschoolde vaklieden, die bovendien een speciale training voor het geven van instructie hebben ontvangen.

In het verleden werd iedere student aan het eind van elke praktikum-middag door zijn instructeur beoordeeld op vier kenmerken: ijver, interesse, volledigheid van het geleverde werk, en de kwaliteit daarvan. Het bleek echter dat het voor de instructeurs vrijwel onmogelijk was op deze wijze tussen de studenten te differentiëren: de standaarddeviatie in hun beoordelingen, die op een tien-puntenschaal werden gegeven, waren voor drie instructeurs resp. 0,24 0,04 en 0,48 (cf. Meuwese en Crombag, 1965). Zelfs de best differentiërende instructeur heeft dus slechts een standaarddeviatie in zijn beoordelingen van een half punt.

De veronderstelling leek gerechtvaardigd dat de gebrekkige differentiatie in de oordelen voor een belangrijk deel het gevolg was van het feit dat de beoordelingscategorieën zo globaal waren. Daarom werd besloten de bij de student te beoordelen gedragingen precieser en vooral concreter te definiëren, opdat het beoordelingsinstrument meer op een observatie-schema zou gaan lijken.

Methode

Door de praktikum-leiding werden, in samenwerking met de instructeurs, een groot aantal gedragsbeschrijvingen verzameld die een "goede praktikant" zou moeten vertonen of integendeel niet zou moeten vertonen.

1) Onderzoek uitgevoerd in opdracht van de Groep Werkplaatstechniek (Prof.ir. C. de Beer), met medewerking van Ir. E. Zweekhorst, B. van Bronckhorst en de instructeurs van het praktikum werkplaatstechniek.

Van deze items leken 52 geschikt voor nader onderzoek. Deze items werden door 27 beoordelaars die allen op de een of andere wijze bij het praktikum betrokken waren (hoogleraren, stafleden, instructeurs) beoordeeld op de mate waarin ze karakteristiek voor "goede praktikanten" geacht werden, te rangschikken op een 7-puntenschaal, lopend van "typerend voor een zeer slechte praktikant" naar "typerend voor de student die alles uit het praktikum haalt wat erin zit". De aldus verkregen beoordelingen werden verwerkt volgens Thurstone's wet van het categorische oordeel (Torgerson, 1958)²⁾. Voor ieder item werd een schaalwaarde verkregen. Tevens werden de grenzen van de schaalclassen berekend. Het middelpunt van de middelste klasse werd bepaald en de schaalwaarden van de items werden omgezet in afwijkingscores ten opzichte van dit middelpunt. De 22 items met de grootste absolute afwijkingscores die bovendien een kleine variantie in de oordelen van de beoordelaars vertoonden, werden geselecteerd. De definitieve lijst bestond uit 12 positieve en 10 negatieve items (zie appendix).

Op deze wijze werd dus de "goede praktikant" gedefinieerd: de student op wie de 12 positieve gedragsbeschrijvingen van toepassing zijn en op wie de 10 negatieve gedragsbeschrijvingen niet van toepassing zijn. Aan de hand van de lijst kan men samenvattend zeggen dat de beoordelaars als een "goede praktikant" defenieerden een student die aandachtig luistert, regelmatig aan de discussie deelneemt, overzichtelijke schetsen en notities maakt en die niet ieder moment hulp van anderen nodig heeft. Van belang is op te merken dat de "goede praktikant" dus niet zozeer als de intelligente dan wel als de ijverige student wordt omschreven.

De lijst werd bij wijze van proef door drie instructeurs gedurende 10 praktikum-middagen gebruikt. Van deze 10 middagen waren er 4 gewijd aan het draaien, waarvan de eerste twee meer theoretisch waren (draaien 1), en de laatste twee overwegend praktisch (draaien 2). Telkens twee middagen waren gewijd aan boren, freeze en schaven. Alle studenten begonnen met de instructie draaien. De volgorde van de volgende onderdelen was niet voor alle studenten hetzelfde. Ook hield een student niet dezelfde instructeur gedurende het hele practikum.

2) De berekeningen werden uitgevoerd door L. van Rooijen en P. de Groot.

Bij het beoordelen tijdens de 6 eerste praktikum-middagen gebruikten de instructeurs de lijst zo, dat zij, wanneer zij een item op de student van toepassing achtten, een teken achter die uitspraak zetten en wanneer zij een item niet van toepassing achtten geen teken zetten. De totaalscore per student werd dan berekend door het optellen van de aangestreepte positieve items en de niet-aangestreepte negatieve items en dit te verminderen met de niet-aangestreepte positieve en de aangestreepte negatieve items. De aldus gescoorde lijst zal verder lijst A worden genoemd.

Het bleek dat het aantal malen dat een item, over de hele groep studenten genomen, aangestreept werd duidelijk afnam naarmate het item verder naar beneden op de lijst voorkwam. Dit doet vermoeden dat een deel van de items helemaal niet gebruikt werd door de beoordelaars.

Na de zesde praktikum-middag werd het formulier zo veranderd dat de instructeur bij ieder item één uit vier antwoorden moest kiezen: "waar", "niet waar", "n.v.t." (niet van toepassing voor dit onderdeel van het praktikum), en "?" (ik weet het niet). Op deze wijze werden de beoordelaars dus gedwongen bij ieder item een oordeel uit te spreken. Deze lijst wordt verder lijst B genoemd. Lijst B werd gedurende de laatste vier middagen van het praktikum gebruikt.

Resultaten

Beoordelingsniveau en spreiding

Zowel bij gebruik van lijst A als van lijst B bleek, dat (a) verschillende instructeurs een verschillend gemiddeld niveau van oordelen vertoonden, en (b) dezelfde instructeur bij verschillende onderdelen van het praktikum eveneens een verschillend gemiddeld oordeelsniveau vertoonde. Dat wordt voor wat betreft lijst B gedemonstreerd in tabel 1.

De verschillen in gemiddeld oordeelsniveau voor de verschillende instructeurs zijn voor alle praktikum-onderdelen significant. Bij het onderdeel schaven blijkt bovendien dat instructeur B meer spreiding in zijn beoordelingen weet te brengen dan instructeur C. Wij kunnen dus concluderen dat er tussen de instructeurs verschillen in "leniency" (mildheid) bestaan, en bovendien dat instructeurs kunnen verschillen in de mate waarin zij verschillen tussen de studenten zien, of wellicht heten verschillen kunnen of willen weergeven met

behulp van de lijst.

Tabel 1¹⁾

Gemiddelde beoordelingen met behulp van
lijst B per onderdeel per instructeur

		Instructeur			z-waarde van het verschil	F-ratio
		A	B	C		
Onderdeel	boren	N 41 \bar{X} 9,61 S 3,12	N 40 \bar{X} 7,23 S 3,67		3,01**	1,18
	frezen	N 45 \bar{X} 6,82 S 2,49		N 53 \bar{X} 9,34 S 2,28	5,14**	1,19
	schaven		N 48 \bar{X} 7,87 S 4,15	N 53 \bar{X} 17,40 S 2,05	14,22**	4,08**
z-waarde van het verschil		4,29**	0,75	19,19**		
F-ratio		1,84*	1,27	1,23		

* $p < .02$; ** $p < .01$ (tweezijdig)

1) N = aantal beoordelingen; \bar{X} = gemiddelde totaalscore; S = standaarddeviatie van de totaalscores.

Bovendien blijkt dat bij twee van de drie instructeurs het gemiddeld oordeelsniveau significant verschilt bij verschillende praktikum-onderdelen: instructeur A kent gemiddeld meer punten toe bij boren dan bij frezen; instructeur C kent gemiddeld meer punten toe bij schaven dan bij frezen. Daar komt nog bij dat instructeur A bij het boren een grotere spreiding in zijn beoordelingen weet te brengen dan bij het frezen.

Samenvattend kan men dus concluderen dat er in de "ruwe" totaalscores een significant instructeurs- en onderdeel- effect zit. Deze effecten maken de verschillende beoordelingen per student onvergelykbaar.

Daarom werden de totaalscores gecorrigeerd voor verschillen in gemiddeld niveau en spreiding per onderdeel en per instructeur. Alle afwijkingsscores werden vervolgens getransformeerd volgens de formule $y = 2x+5$, waarbij x de afwijkingsscore is en y de getransformeerde score. Op deze wijze werd het gemiddelde der gestandaardiseerde scores gelijk aan 5. De totale variatie-breedte bedroeg ca. 11 punten. Bij het standaardiseren werden lijst A en lijst B gescheiden gehouden. Het bleek bij lijst A dat de standaarddeviatie van de beoordelingen nu 3,96 bedroeg, hetgeen een aanmerkelijke vooruitgang betekent in vergelijking met de vroeger gebruikte methode. Bij lijst B bleek de spreiding nog groter geworden, namelijk 4,75. Op een schaal van circa 11 punten wordt nu dus een standaard-deviatie van bijna 5 punten gevonden, terwijl vroeger op een 10-puntenschaal een standaard-deviatie van nog geen half punt werd gevonden. We kunnen dus concluderen dat de nieuwontworpen beoordelingslijst het de instructeur mogelijk maakt beter tussen de studenten te discrimineren. Dwingt men door geprecodeerde antwoorden de beoordelaars werkelijk alle items te gebruiken dan gaat de spreiding nogmaals omhoog. Het verschil tussen de standaard-deviatie van lijst A en die van de lijst B is nagenoeg significant ($F = 1,51, p .05$)

Betrouwbaarheid

De betrouwbaarheid van het beoordelingsinstrument werd bepaald met behulp van de "split-half" methode (even versus oneven items). Voordat echter de sub-totalen met elkaar gecorreleerd konden worden moesten ook deze gecorrigeerd worden voor instructeurs- en onderdeel-effecten. De sub-totalen werden dus op dezelfde wijze bewerkt als eerder de totaalscores. Deze bewerking werd uitgevoerd op de beoordelingen uitgebracht op de eerste twee praktikum-middagen dat lijst B gebruikt werd (de 7^e en de 8^e week³⁾).

De betrouwbaarheid van de beoordelingen uitgebracht in de zevende week bleek 0,79 te zijn, die van de beoordelingen over de achtste week 0,78.

3) De bewerking werd tot deze twee weken beperkt, omdat daarin onderdeel en instructeur gecontamineerd waren. De correcties konden zo op een voldoende aantal studenten berekend worden.

Dit resultaat mag beschouwd worden als bevredigend, zowel wat betreft de grootte van de getallen als wat betreft de stabiliteit ervan.

Validiteit

Naast de scores voor het praktikum waren van de studenten nog een reeks andere gegevens bekend. Daaruit werden per student 10 gegevens geselecteerd die op het eerste gezicht wellicht zouden kunnen meespelen in de beoordeling die de instructeur van de student geeft. Deze gegevens waren: (1) intelligentie (Technische Hogeschool Delft, 1959); (2) studie-motivatie (Meuwese en Crombag, 1964); (3) studie-attitude (Meuwese en Crombag, 1964); (4) de "LPC"-score: "esteem for the least preferred co-worker" (Fiedler, 1962); (5) urbanisatiegraad van de plaats van herkomst (een modificatie van de C.B.S.-codering); (6) milieu (eveneens een modificatie van de C.B.S.-codering); (7) cijfer voor het wiskunde 1-tentamen; (8) cijfer voor het mechanica 1-tentamen; (9) een score voor aanpassing aan het universitair milieu (vragenlijst waarvoor afzonderlijk zal worden gerapporteerd; en (10) leeftijd. Deze 10 scores werden als predictoren gehanteerd.

De beoordelingen voor het praktikum werden als criteria beschouwd. Aparte scores werden berekend voor de twee eerste middagen van de instructie in het draaien, die een meer theoretisch karakter hadden (kinematica), en de twee laatste middagen van het draaien, die meer praktisch van aard waren. Zo ontstonden dus twee scores voor het draaien: draaien 1 en draaien 2. Voorts waren er nog afzonderlijke scores voor boren, frezen en schaven. Van de laatste drie scores is het onderdeel boren wat meer theoretisch van aard dan frezen en schaven. Omdat boren wat de praktische uitvoering ervan relatief de eenvoudigste bewerking is, werden in de middagen die bestemd waren voor het boren meer theoretische beschouwingen opgenomen.

Er zijn dus 5 criteria.

Met deze 10 predictoren en 5 criteria werd een canonische correlatie analyse uitgevoerd⁴⁾ (Cooley & Lohnes, 1962). Bij een dergelijke analyse wordt geprobeerd zowel de predictoren als de criteria zodanig te wegen en lineair te combineren dat de correlatie tussen beide maximaal is. De canonische correlatie kan worden beschouwd als een index voor

de mate waarin individuen dezelfde relatieve posities innemen in de p-dimensionale ruimte van de predictoren als zij innemen in de q-dimensionale ruimte van de criteria.

Er werd één significante canonische correlatie gevonden: zie tabel 2. In deze tabel zijn tevens voor de predictoren en de criteria de gewichten weergegeven. Deze gewichten geven dus aan op welke wijze men zowel predictoren als criteria moet wegen om tot een correlatie van 0,64 tussen beide te komen.

Tabel 2

Canonische correlatie en gewichten praktikanten-beoordeling

Canonische correlatie .64 $\chi^2 = 66,35$ $df = 50$ $.10 > p > .05$

Gewichten criteria:

draaien 1 -.48
draaien 2 .40
boren .02
frezen .33
schaven .71

Gewichten predictoren:

intelligentie .03
studiemotivatie .47
studie-attitude .16
l.p.c. .46
urbanisatiegraad-.10
milieu .20
wiskunde 1 -.08
mechanica .53
aanpassing -.04
leeftijd -.46

De criteria blijken zodanig gewogen te zijn dat de scores voor de meer praktische praktikum-onderdelen een hoog gewicht krijgen. Van de meer theoretische onderdelen dient draaien 1 als suppressor; het boren krijgt een gewicht van 0. Bij de predictoren krijgen studiemotivatie, LPC en het mechanicacijfer de hoogste gewichten, terwijl leeftijd een suppressor is. Het feit dat het wiskundecijfer een gewicht van nul krijgt behoeft niet zonder meer significant te zijn, omdat mechanica en wiskunde in de onderzochte groep gecorreleerd zijn ($r = 0,47$). De overige intercorrelatie tussen de predictoren zijn alle belangrijk lager.

Geconcludeerd kan worden dat hoge scores voor de praktische onderdelen vooral behaald worden door studenten met een hoge studie-motivatie (de ijverigen), met een hoge LPC-score (de wat meer tolerante studenten, die zelfs iemand met wie zij helemaal niet kunnen samenwerken nog relatief gunstig beoordelen), door studenten met hoge cijfers voor mechanica en door jongere studenten. De hoge gewichten voor motivatie en mechanica pleiten voor de validiteit van de beoordelingen. De hoge gewichten voor LPC en leeftijd (in deze steekproef ongecorrigeerd met wiskunde en mechanica) demonstreren dat ook andere, minder relevant te achten factoren in het oordeel van de instructeur kunnen meespelen. Bij het bovenstaande moet men er echter rekening mee houden, dat de gebruikte analyse-methode erg gevoelig is voor toevalligheden in de steekproef.

Toepassing

Bij het gebruik van het instrument kan men het best per student de scores over het gehele praktikum combineren. Het is nodig daarbij de scores per instructeur per onderdeel te corrigeren voor gemiddeld niveau en voor verschillen in spreiding (standaardiseren). Bij verschillen in spreiding zou anders de instructeur met de grootste spreiding in zijn beoordelingen relatief de grootste invloed op de gecombineerde totaalscores gaan uitoefenen.

De per onderdeel per instructeur gestandaardiseerde scores kan men per student middelen. Aangenomen dat de betrouwbaarheid van de scores voor alle middagen in dezelfde orde van grootte ligt als de voor de zevende en achtste week gevonden betrouwbaarheden, dan zal de betrouwbaarheid van de gecombineerde score hoger zijn dan die van de afzonderlijke scores.

De totaalscore kan men tenslotte het best transformeren naar een 10-punten schaal. Dit kan op vele manieren gebeuren, bijvoorbeeld door in oorspronkelijke verdeling op willekeurige punten grenzen aan te brengen. Het verdient echter aanbeveling de intervallschaaleigenschappen van de scores te handhaven. In dat geval is alleen een lineaire transformatie van de scores naar een 10-punten schaal toegestaan.

De volgende procedure kan worden aanbevolen.

1. Bepaal de minimaal voldoende score x_1 in de oorspronkelijke verdeling; dit moet op de 10-punten schaal $y_1 = 6$ worden.

2. Bepaal de score x_2 in de oorspronkelijke verdeling waaraan het oordeel "uitmuntend" wordt toegekend; dit moet op de 10-punten schaal $y_2 = 10$ worden. Deze score kan al dan niet in de werkelijk gevonden verdeling voorkomen.

3. Los a en b op uit de vergelijkingen:

$$y_1 = ax_1 + b$$

$$y_2 = ax_2 + b$$

4. Bereken voor elk individu het aantal punten y_i uit de score x_i als volgt:

$$y_i = ax_i + b$$

5. Rond de punten af op hele getallen. Rond cijfers lager dan 1 af op 1; rond cijfers tussen 5 en 6 af op 5.

Discussie

Gebleken is dat door het concreter maken van de door de instructeurs te beoordelen gedragingen de mate van differentiatie die zij onder de studenten weten aan te brengen zeer sterk toeneemt. Tegelijk daarmee zijn ongetwijfeld ook de betrouwbaarheid en de validiteit van de scores toegenomen. Immers, scores die niet of nauwelijks spreiden schieten noodzakelijk ook tekort in betrouwbaarheid en validiteit.

De betrouwbaarheid van het beoordelingsinstrument is bevredigend gebleken. De validiteit van het instrument is in eerste instantie een kwestie van definitie. De praktikum-leiding heeft gedefinieerd (met behulp van de schaalprocedure) dat ieder op wie de 12 positieve beschrijvingen van toepassing zijn en op wie de 10 negatieve beschrijvingen niet van toepassing zijn een "goede praktikant" is. Wanneer dit van toepassing zijn van de gedragsbeschrijvingen nu maar op betrouwbare wijze geschiedt - en dat is zo - dan is ieder die een hoge score haalt een goede praktikant, zoals gedefinieerd door de items.

Al eerder is opgemerkt dat de leiding een goede praktikant niet zozeer als een intelligente dan wel als een ijverige student gedefinieerd heeft. De onder het hoofd "validiteit" gepresenteerde resultaten van de canonische correlatie analyse bevestigen dit.

Samenvatting

Ter beoordeling van praktikanen in de werkplaatstechniek werd een beoordelingsinstrument geconstrueerd, bestaande uit een reeks van 22 concrete gedragsbeschrijvingen.

De lijst werd gedurende 10 praktikum-middagen bij wijze van proef gebruikt. Met behulp van de lijst bleken de beoordelaars goed tussen de praktikanten te kunnen differentiëren. Het instrument bleek voldoende betrouwbaar. Argumenten voor de validiteit van het instrument werden besproken.

Referenties

Cooley, W.W., & P.R. Lohnes

Multivariate procedures for the behavioral sciences
New York, Wiley, 1962

Fiedler, F.E.

Leader attitudes, group climate, and group creativity
J. abnorm. soc. Psychol, 1962, 65, 308-318

Meuwese, W. & H. Crombag

Constructie van instrumenten voor de meting van studiemotivatie en studie-attitude
Rapport nr. 5, Groep Onderwijsresearch
Eindhoven, Technische Hogeschool, 1964.

Meuwese, W. & H. Crombag

Een proefneming met beoordelingen van instructeurs door studenten
Tijdschrift. Psychol. Gawein, 1965, 13, 109-120

Technische Hogeschool Delft

Mislukking en vertraging van de studie
Delft, 1959

Torgerson, W.S.

Theory and methods of scaling
New York, Wiley, 1958

BEOORDELINGSLIJST.

1. Hij luistert aandachtig.
2. Hij neemt regelmatig deel aan de discussie.
3. Hij volgt aanwijzingen goed op.
4. Hij maakt duidelijke schetsen.
5. Hij kijkt af bij het schetsen.
6. Hij weet vlot op vragen te antwoorden.
7. Hij heeft vlot een eigen oplossing bij de hand.
8. Hij maakt storende opmerkingen.
9. Hij maakt nette en overzichtelijke schetsen.
10. Hij knijpt er vroeger tussen uit.
11. Hij laat zich door anderen helpen bij zijn werkstuk.
12. De instructeur moet hem voortdurend helpen.
13. Zijn aantekeningen zijn overzichtelijk.
14. Zijn schetsen hebben de goede verhoudingen.
15. Bij groepswork neemt hij de leiding.
16. Bij groepswork laat hij de anderen alles opknappen.
17. Hij zit te niks als een bepaalde bewerking gedaan is.
18. Zelfs eenvoudige handelingen moeten hem dikwijls worden voorgedaan.
19. Hij kan zijn ideeën goed motiveren.
20. Hij was zijn spullen vergeten.
21. Hij prutst maar wat aan.
22. Hij stelt vragen naar aanleiding van de vorige middag.