

NOKUT-portalen som inngangsport i kvalitetsarbeid: utfordringer ved karakterer som kvalitetsindikator

Tilsynsinngangen i NOKUT-portalen muliggjør enkle, kvantitative analyser, som kan tjene som inngangsport for mer målrettede undersøkelser av kvaliteten i norske utdanningsinstitusjoner og -programmer. Dette mulighetspotensialet er stort, både for NOKUTs kvalitetsarbeid og for institusjonenes eget kvalitetsarbeid. Aktuelle analyser vil i en serie artikler fremover vise eksempler på hvordan NOKUT-portalen kan anvendes, og diskutere bruken av indikatorene i portalen med utgangspunkt i relevant forskning. Denne artikkelen tar for seg et par indikatorer fra studentdata. Vi begynner med et tema som er velkjent for de fleste innen kvalitetsarbeid, nemlig sammenhengen mellom karakterpoeng og eksamensresultater som indikator på læringsutbytte mellom institusjoner.

Bakgrunn

Kvantitative data er både relevant og nødvendig i arbeidet NOKUT er satt til å gjøre gjennom «tilsynsforordningen». Indikatorer som dekker ulike aspekter ved utdanningskvalitet er utledet av forordningen og gjort tilgjengelig i NOKUT-portalen. Som dokumentasjonen av indikatorene påpeker, gir ikke den enkelte indikator alene et godt bilde av egenskaper ved kvaliteten på institusjonen/studieprogrammet. Sammen kan flere indikatorer imidlertid skape et bilde av situasjonen, og slik tjene som bakgrunnsinformasjon og som utgangspunkt for videre kvalitetsarbeid. Gjennom bruk av NOKUT-portalen kan vi altså reflektere over resultater som foreligger på et overordnet nivå, og bruke dette som et utgangspunkt for mer detaljerte undersøkelser av utdanningskvalitet. Dette er relevant både for institusjonenes eget kvalitetsarbeid, og i NOKUTs tilsynsarbeid. Relevansen for sistnevnte kan også leses av «Forslag om nasjonal strategi for utdanningskvalitet og ny modell for tilsyn med høyere utdanning» (NOKUTs innspill til Kunnskapsdepartementet desember 2012)¹. I forslaget anbefaler NOKUT å effektivisere tilsynsprosessen ved at vår egen administrasjon gjør mindre bruk av sakkyndige i begynnelsen av tilsynsprosessen, men selv innleder med risikoanalyser og forberedelser basert på grunnlagsdata i nasjonale baser. NOKUT-portalen utgjør i denne sammenheng en kilde for slike grunnlagsdata.

Det er bred enighet om at vi trenger mer kunnskap om kvalitetsforskjeller mellom de enkelte universiteter og høyskoler, og Kunnskapsdepartementet har spesielt påpekt at det finnes lite kunnskap om hvorvidt og i hvilken grad karakterskalaen brukes enhetlig mellom utdanningsinstitusjonene. Bruk av karakterskalaen er relevant i vurderingen av studenters læringsutbytte og utdannings merverdi (på engelsk kjent som «value added»), og er derfor i utgangspunktet en interessant indikator på kvalitet i utdanningen. Dersom læresteder bruker karakterskalaen forskjellig, hefter det imidlertid utfordringer ved kvalitetsvurderinger basert på avgangskarakterer. En måte å undersøke institusjonenes bruk av karakterskalaen, er ved å se på

1

http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Generell/Om%20NOKUT/Forslag_om_nasjonal_strategi_for_utdanningskvalitet_og_ny_modell_for_tilsyn_med_h%c3%b8yereutdanning_2012.pdf

avgangskarakterer justert for inntakskarakterer. NOKUT-portalen gir i denne sammenheng mulighet for å sammenlikne studentenes inntakskarakterer med andelen studenter som får A eller B, og kan dermed indikere forskjellig bruk av karakterskalaen. Dette kan tjene som en inngangsport til videre studier av utdanningskvalitet, læringsutbytte og karaktersetting. Dersom inntakskarakternivået er relativt lavt, men andelen studenter som oppnår de to beste karakterene er høyt, kan det indikere godt læringsutbytte og høy utdanningskvalitet, men det kan også reflektere «snill» bruk av toppnivåene i karakterskalaen. Motsatt, om studentenes karaktarnivå ved inntak er høyt, mens andelen som oppnår de to beste karakterene på eksamen er lav, kan det reflektere manglende læringsutbytte og studiekvalitet – men kan hende er det mer nærliggende å anta at det handler om «streng» bruk av toppnivåene i karakterskalaen. I denne artikkelen skal vi bruke NOKUT-portalen til å se på hvordan statlige høyskoler og universiteter fordeler seg etter indikatorene «karakterpoeng blant nye studenter» og «andel eksamenskarakterer A + B». Ettersom de to indikatorene representerer forskjellige skalaer – henholdsvis poeng og andel/prosent, gir det ikke mening å se på differansen mellom de to indikatorene, men vi kan foreta en kvalitativ vurdering. For å kunne diskutere resultatet mer detaljert, begrenser vi analysen til lærerutdanninger og yrkesutdanninger i pedagogikk, med undergrupperingen fag- og yrkesfaglærerutdanninger. Resultatene av analysen og utfordringer relatert til potensielle videre undersøkelser av læringsutbytte, diskuteres i relasjon til relevant forskning.

Analysen

Tabell 1 viser karakterpoeng blant nye studenter, samt strykprosent og andel eksamenskarakterer A og B, blant universiteter og statlige høyskoler i NOKUT-portalen. Som det fremgår av tabellen, ser vi kun på faggruppen «Fag- og yrkesfaglærerutdanninger», og begrenser analysen til utdanninger på bachelornivå.

Karakterpoeng

Tabellen viser at inntakskvaliteten er temmelig lik totalt sett blant de to institusjonsformene, med universiteter (38,6) knapt ett karakterpoeng høyere enn statlige høyskoler (37,8). Det er imidlertid større spredning på inntakskvaliteten mellom enkeltinstitusjoner enn mellom de to institusjonsformene. Blant høyskolene varierer antallet karakterpoeng fra 35,5 til 42,3 – noe som gir en variasjonsbredde på hele 6,8 poeng – mens ved universitetene varierer det fra 37,4 til 40,3 – en variasjonsbredde på kun 2,9 poeng. Variasjonen i inntakskvaliteten er altså større blant høyskoler enn blant universiteter.

Strykprosent

Resultatene fra strykprosent-analysen i tabellen, viser i grove trekk samme tendens som karakterpoeng-analysen. Variasjonsbredden for høyskolene er på 4,2 prosent, mens den for universitetene er på 3 prosent; det er altså igjen noe større spredning blant høyskolene. I snitt ligger strykprosenten på 3,5 blant høyskolene, og 2,2 blant universitetene – igjen om lag ett prosentpoeng i favør av universitetene – men det er store variasjoner innad blant de to institusjonsnivåene. NTNU skiller seg markant ut ved å ha en strykprosent på kun 0,7, trass i at de har det laveste antall karakterpoeng blant nye studenter, når vi kun ser på data på universitetsnivå. Til sammenlikning har Høyskolen i Hedmark – institusjonen blant høyskoler i tabellen som har det laveste antallet karakterpoeng blant nye studenter – også den høyeste strykprosenten. Umiddelbart virker dette som et mer realistisk resultat med tanke på utdanningens merverdi sett i relasjon til

inntaksvalitet, men det vanskelig å si hva som ligger bak tallene, uten å ty til spekulasjoner. Her kunne det vært interessant å gjøre noen videre analyser på hva som skiller de to institusjonenes utdanningsopplegg ved Fag- og yrkesfaglærerutdanninger.

Andel A og B

Den siste kolonnen i tabellen viser resultater på indikatoren Andel eksamenskarakterer A+B. Når det gjelder variasjonsbredde i resultatene, skiller denne analysen seg fra de to øvrige ved at variasjonsbredden er større blant universitetene enn blant høgskolene: 35,4 prosentpoeng kontra 12,6 prosentpoeng. Ved Universitetet i Agder, der antallet karakterpoeng blant nye studenter ligger på 37,8, er andelen A og B nesten 70 prosent. Tatt i betraktning det relativt gjennomsnittlige nivået på antall karakterpoeng i denne høgskolens studentgrunnlag, er andelen som kommer ut med toppkarakterer høyt. Til sammenlikning er andelen A og B «kun» på om lag 50 prosent ved Høgskolen i Nord-Trøndelag, til tross for at antallet karakterpoeng blant nye studenter er høyere enn ved Universitetet i Agder. Igjen er det umulig å si noe om hva forskjeller som disse skyldes kun basert på tallene i NOKUT-portalene. Resultatene kan imidlertid fungere som grunnlag for mer detaljerte undersøkelser av kvaliteten på studieprogrammene i den aktuelle faggruppen.

Tabell 1: Karakterpoeng blant nye studenter, strykporsent og andel eksamenskarakterer A + B blant statlige universiteter og høyskoler

	Indikator: Karakterpoeng blant nye studenter* Fagfelt: Lærerutdanninger og utdanninger i pedagogikk Faggruppe: Fag- og yrkesfaglærerutdanninger	Indikator: Strykporsent Fagfelt: Lærerutdanninger og utdanninger i pedagogikk Faggruppe: Fag- og yrkesfaglærerutdanninger	Indikator: Andel eksamenskarakterer A + B** Fagfelt: Lærerutdanninger og utdanninger i pedagogikk Faggruppe: Fag- og yrkesfaglærerutdanninger
Statlige høyskoler			
Høgskolen i Akershus	36,4	1,6	50,2
Høgskolen i Hedmark	35,5	7,4	38,1
Høgskolen i Nord-Trøndelag	40,1	2,3	50,7
Høgskolen i Oslo	42,3	5,8	41,4
Høgskolen i Telemark	35,6	4,8	43,9
Høgskolen Stord/Haugesund	36,8	2,5	43,7
Gjennomsnitt, statlige høyskoler***	37,8	3,5	44,7
Universiteter			
Norges teknisk-naturvitenskapelige universitet	37,4	0,7	62,1
Universitetet i Agder	37,8	2,5	69,7
Universitetet i Bergen	40,3	2,5	34,3
Universitetet i Nordland	39,0	3,7	34,5
Gjennomsnitt, universiteter****	38,6	2,2	50,2

Note: Alle tall er for 2011. Kun institusjoner som har tall på samtlige tre indikatorer i NOKUT-portalene er med i oversikten over enkeltinstitusjoner. For studentdata er kun programmer på nivå «Bachelor» tatt med. Nivåkoder: B3, B4.

* Indikatoren er relevant som bakgrunnsinformasjon i NOKUTs evalueringer av kvalitetssikring i utdanningen;

Indikatoren er relevant grunnlag for å undersøke hvorvidt institusjonenes kvalitetssikringssystem fanger opp årsaker til dette (altså, årsaker til andelen A'er og B'er). * Snittet inkluderer samtlige høyskoler som det finnes tall på i NOKUT-portalene **** Snittet inkluderer samtlige universiteter som det finnes tall på i NOKUT-portalene

For ordens skyld bør det tilføyes at det ikke er noe i veien med at institusjoner eller utdanningsprogram anvender seg av noe *ulike deler* av karakterskalaen, så lenge karakterene som settes er *like i henhold til kriteriet* som er fastsatt i Lov om universiteter og høyskoler § 3-9. Karakterskalaene i norsk høyere utdanning er nemlig kriteriebasert og formulert som absolutte – det vil si at karakterene ikke skal fastsettes i forhold til en gitt fordeling («normalfordeling»), men at en bestemt karakter skal gis så lenge prestasjonen tilfredsstillende kriteriene for denne karakteren. Det følger av dette at andel stryk og A og B på eksamen kan variere betydelig mellom læresteder og utdanningsprogram, uten at det kan tas til inntekt for forskjellig bruk av karakterskalaen. Dersom inntakskvaliteten på studentmassen er relativt lav, men resultat kvaliteten er høy, er det imidlertid grunn til å undersøke nærmere hva som ligger bak. Det er interessant for å avdekke forskjeller i bruk av karakterskalaen, og ikke minst for å undersøke om noen typer undervisningsopplegg gir bedre resultat for studentenes læring enn andre typer.

Oppsummering av analysen

Hittil har vi sett på noen indikatorer fra studentdata i NOKUT-portalen. Ved å sammenlikne utdanningsinstitusjoner langs de tre indikatorene, har vi lagt merke til at det ikke nødvendigvis er samsvar mellom inntakskvalitet og resultat kvaliteten hos studentgrunnet. Mer presist har vi sett at det er store variasjoner mellom lærestedene, særlig når vi ser karakterpoeng blant nye studenter opp mot andel A og B på eksamen. Det finnes altså forskjeller i karakterfordelinger som ikke kan tilskrives forskjeller i opptaksgrunnet. Slike forskjeller kan i hovedsak tolkes på to måter: Enten som et resultat av variasjoner i bruken av karakterskalaen ved lærestedene, eller som et resultat av kvalitetsforskjeller ved lærestedenes studieprogram og undervisning (i dette eksempelet, fagfeltet «Fag- og yrkesfaglærerutdanninger» på bachelornivå). Det er selvsagt også mulig at begge disse tendensene opptrer parallelt. I og med dette store spriket i fortolkningsmuligheter, har forskere spurt seg om det kan festes lit til karakterer som mål på utdanningens merverdi (læringsutbytte). De følgende to avsnittene vil ta for seg noen utfordringer ved karakterer som kvalitetsindikator, og skissere potensielle forklaringer på forskjeller i resultat kvaliteten, som kan fungere som utgangspunkt for videre analyser. For ryddighetens skyld, ser vi først på problemstillinger tilknyttet såkalt «objektive» data – slike som NOKUT-portalen gir tall på – og går deretter over til mer subjektive data, det vil si intervjudata fra spørreskjema eller kvalitative intervjuetknikker.

Karakterer som mål på utdanningskvalitet

Det er tidligere påvist at det er ulik bruk av karakterskalaen mellom ulike læresteder og mellom fagområder. I norsk sammenheng har Møen & Tjelta (2005) gjort analyser som viser at det var systematiske skjevheter i bruk av karakterskalaen mellom ulike statlige høyskoler med økonomisk-administrative studier i perioden 1998-2003. Ved å sammenlikne studenter ved Norges Handelshøyskole (NHH) med 529 studenter fra 23 høyskoler, fant de at studenter som fortsatte sine studier på NHH etter å ha påbegynt utdanningen sin ved en annen høyskole, ofte falt dramatisk i karakternivå, trass i at de hadde høye karakterer fra sitt første lærested. Selv om det var positiv sammenheng mellom karakterer fra tidligere lærested og prestasjoner ved NHH, var det store forskjeller ut i fra hvilket lærested studenten kom fra. Til tross for at det norske karaktersystemet er kriteriebasert, og trass i at Universitets og høyskolerådet (UHR) legger ned betydelig innsats for å sikre at karaktersystemet anvendes likt på tvers av læresteder og utdanningsprogrammer² tyder altså noe på at det av og til praktiseres på en normbasert måte. Normbasert karaktersetting betyr at en prestasjon vurderes i forhold til de andre prestasjonene i gruppen, det vil si at karaktersettingen er relativ i forhold til gruppen av studenter som sensureres. Over tid vil slike karakterer fordeles normalt i alle fag, noe som betyr at det er problematisk å benytte karakterene til å analysere kvalitetsforskjeller. Det trenger ikke ligge bevisst handling bak forskjeller i karaktersetting. For eksempel kan karaktersetting tenkes å variere mer blant «myke» enn «harde» fagdisipliner, fordi det

² <http://www.uhr.no/index.php?objectId=3899&method=contents>

er enklere å skille mellom riktige og gale svar i de «harde» disiplinene. Også forskjeller i eksamensordninger (slik som mappe-evalueringer; gruppeoppgaver; muntlig og skriftlig eksamen m.m.) kan tenkes å påvirke karakterfordelingen. Det ligger også en rekke forutsetninger til grunn knyttet til vurderingen av den enkelte students prestasjoner (Lauvås, Havnes, & Raaheim, 2000), slik som standardisering av eksamenssituasjonen, objektivitet i bedømmingen, bruk av eksterne sensorer som «rettesnor» for bedømmingen, samt at bedømmingen dekker de læringsresultater utdanningen er formulert å dekke – verken mer eller mindre. I tillegg vil det være forskjeller i hvordan ulike faglige forhold vektlegges innenfor samme fag som undervises på forskjellige steder. Ordninger for å vurdere en individuell prestasjon er altså svært sammensatt, og det er derfor ikke uproblematisk å aggregere karakterdata på individnivå opp til studium- eller lærestednivå.

Utfordringer til tross, eksamensresultater er relevante kvalitetsindikatorer. Dersom det finnes en tendens til at forskjellen mellom inntakskarakterer og avgangskarakterer er størst i de tilfeller hvor inntakskarakternivået er lavest, er det nærliggende å tenke seg at karaktersetningen er relativ til inntakskarakterer. For eksempel, dersom institusjoner med forskjellig søkermasse hva angår karaktersnitt fra videregående, utviser relativt like karakterfordelinger ved uteksaminering av sine kandidater, kan det bety at noen institusjoner er «snillere» enn andre i karaktersetningen. Det å begynne kvalitetsanalyser med store grupper og fra et bredt perspektiv, er altså en relevant måte å innlede mer detaljerte analyser av kvalitet på lavere nivåer i utdanningsystemet.

Studentrapportert læringsutbytte som mål på utdanningskvalitet

Med tanke på alle måleutfordringene som knytter seg til karakterdata, er det nærliggende å ty til andre datakilder for å måle kvaliteten i høyere utdanning. Dette har også blitt gjort, dog ikke først og fremst fordi registerdata er svake, men fordi surveydata ofte er et billigere alternativ. Subjektive, retrospektive tverrsnittsdata er rimeligst, og er derfor den hyppigst fremtredende måten å måle merverdien på i høyere utdanning. Slike data er imidlertid beheftet med sine egne utfordringer, fordi mennesker er lite treffsikre når det gjelder å vurdere hvordan og hvorfor deres kunnskap har endret seg i løpet av en gitt tidsperiode (Bowman 2010). Bowman (2010, 2011) er blant de som har funnet at det nesten ikke er noe samsvar mellom karakterer målt i første og siste år av høyere utdanning og studentrapportert læring målt i siste år av utdanningen. Med andre ord, når objektive og subjektive data på læringsutbytte for en og samme person sammenliknes, indikerer forskningen at personen selv er tilbøyelig til å mene noe helt annet om sitt læringsresultat enn det endringen i personens karakterer gir uttrykk for.

Kuh (2003), som administrerer en nasjonal studentsurvey i USA (NSSE), hevder imidlertid at selvrapporterte opplysninger er sannsynliggjort å være gyldige så lenge fem forutsetninger er oppfylt, mer presist når: (1) informasjonen som etterspørres er kjent for informanten; (2) spørsmålene er formulert klart og utvetydig; (3) spørsmålene omhandler aktiviteter nær i tid; (4) respondentene synes spørsmålet fortjener et seriøst og gjennomtenkt svar; (5) å svare på spørsmålet er ikke pinlig, truende eller bryter privatlivet til respondenten, eller oppfordrer til å svare på en sosialt ønskelig måte. Kuh (2003 s 3-4) understreker dessuten at forskning (Pike, 1995) har vist at studentrapportert læringsutbytte er høyt korrelert med relevante testskårer for prestasjoner, slik som karakterer. Pascarella (2001) deler Kuh's oppfatning om dette, men understreker at problemet med tverrsnittsundersøkelser ikke knytter seg først og fremst til individers (manglende) evne til å vurdere egen læring. Snarere ligger problemet i manglende bruk av kontrollvariabler forut for høyere utdanning, som er gode prediktorer for studentrapportert læringsutbytte, mener han. Mangelen på longitudinelle undersøkelser med relevante pretest (før-test)-variabler, gjør at mange undersøkelser ikke er i stand til å skille effekten av "sosialisering", altså hvordan studentene endrer seg intellektuelt som følge av utdanningen, fra effekten av "rekruttering", det vil si hvilken type studentmasse som søker seg til utdanningsinstitusjonen. Pascarella's (2001) retter denne innvendingen både mot tverrsnittsstudier basert på registerdata og spørreskjemadata.

Astin og Lee (2003) påpeker for øvrig ytterligere en dimensjon ved sosialisering-problematikken, nemlig bruken av tverrsnittundersøkelser for rangering og sammenlikning av høyere utdanningsinstitusjoner. For å vise hvor problematisk dette er, anvender Astin og Lee data fra en studentsurvey for å konstruere gjennomsnittsmål på forventet og faktisk studentinnsats ved 222 høyere utdanningsinstitusjoner. Først stilte forskerne spørsmål om timer brukt på studier i uka i videregående skole (pre-test). Deretter stilte de spørsmål om timer brukt på studier i uka i den høyere utdanningsinstitusjonen etter at studentene har blitt uteksaminert (post-test). Resultatene viste at korrelasjonen mellom svarene studentene ga før de påbegynte høyere utdanning og etter uteksaminering er høy, og at utdanningsinstitusjoner som kom godt ut på rangeringer derfor ikke nødvendigvis kan ta æren for at studentene deres er spesielt engasjerte. Som eksempel viser Astin og Lee til et høyt rangert universitet, der studentene brukte i snitt 15.8 timer i uka på hjemmearbeid/egenstudier. Universitetet kom godt ut på rankingen, men kan resultatet tas som et tegn på at utdanningen studentene har fått ved universitetet er spesielt engasjerende? Nei, det kan de ikke, for det forventede antall timer (basert på data fra pre-testen) studentene skulle brukt på hjemmearbeid var 15.9 timer i uka. Institusjonen har altså ikke gjort annet enn å opprettholde innsatsen studentene fra før av var tilbøyelige til å nedlegge i sine studier. "(The) institution manages to engage its students at precisely the level that would be expected from their entering characteristics", konkluderer Astin og Lee (2003, s 667).

Mer overordnet, i forhold til måling av utdannings merverdi, kan det helt til slutt nevnes at Pascarella et al. (2011) også kritiserer mangelen på gode mål på hva som bør være *den forventede endringen* i kompetanse i løpet av høyere utdanning. Så lenge vi ikke har et slikt mål, er det vanskelig å si hva som skiller institusjoner/utdanninger som er gode nok fra de som ikke er gode nok. I tillegg, påpeker Pascarella et al. (2011) er det ikke utført sammenlikninger av studenters kompetanseendring og endringen i kompetansen til individer som ikke studerer, noe som gjør det umulig å avgjøre hva høyere utdanning *spesifikt* tilfører, ut over kunnskap og ferdigheter man får på andre samfunnsarenaer.

Konklusjon og videre arbeid

Som denne gjennomgangen har vist, ser forskere utfordringer ved design og datatilfang i undersøkelser av prestasjoner i høyere utdanning. Det bør være opplagt for leseren at poenget med en gjennomgang av disse utfordringene ikke å parkere forsøk på å drive kvalitetsarbeid i høyere utdanning, eller å senke ambisjonene om å måle kvalitet på gode måter. Formålet er å bevisstgjøre på utfordringene, og å løfte frem undersøkelser og analyser som kan gjennomføres i etterkant av enkle innledende analyser. Sett under ett er det mye som tyder på at karakterdata er en av de bedre indikatorene vi har på kvalitet i utdanningen når det gjelder vurdering av læringsutbytte og merverdi – ikke fordi de er spesielt gode, men fordi det ikke eksisterer bedre indikatorer. Karakterdata må altså anvendes med et bevisst forhold til måleproblematikken, og oppsiktsvekkende funn bør følges opp med ytterligere undersøkelser. Tall fra NOKUT-portalene muliggjør nettopp innledende analyser av kvalitet i høyere utdanning, og kan avdekke forskjeller mellom institusjoner og utdanninger som det kan være interessant å forfølge videre.

Fremtidige numre av Aktuelle Analyser vil være orientert blant annet mot å utvide bruken av NOKUT-portalene i kvalitetsarbeid internt og eksternt. NSD har ennå ikke utviklet alle indikatorene NOKUT ønsker, men vi forventer å få flere interessante indikatorer – blant annet mål på gjennomstrømning – allerede i løpet av 2013. Disse indikatorene vil bli omtalt i fremtidige numre av Aktuelle Analyser.

Artikkelen er skrevet av Hilde Karlsen

Kilder

- Astin, A. W. & Lee, J. J. (2003). 'How risky are one-shot cross-sectional assessments of undergraduate students?', *Research in higher education* 44: 657-72
- Bowman, N. A. 2010. Can 1st-year college students accurately report their learning and development? *American educational research journal*, 47(466-496).
- Bowman, N. A. 2011. Validity of college self-reported gains at diverse institutions. *Educational researcher*, 40.
- Kuh, G. D. 2001. Assessing what really matters to student learning: Inside the national survey of student engagement. *Change*, 33(10-17).
- Kuh, G. D. (2003). 'The national survey of student engagement: Conceptual framework and overview of psychometric properties': Indiana University Center for Postsecondary Research and Planning, Indiana.
- Lauvås, P., Havnes, A., & Raaheim, A. 2000. Why this inertia in the development of better assessment methods? *Quality in Higher education*, 6(1): 91-99.
- Møen, J., & Tjelta, M. 2005. Bruker ulike høyskoler karakterskalaen ulikt? En analyse av sammenhengen mellom skolebakgrunn og faglig suksess ved NHH. *Økonomisk forum*(6).
- Pascarella, E. 2001. Using student self-reported gains to estimate college impact: A cautionary tale. *Journal of college student development*, 42: 488-492.
- Pascarella, E. T., Blaich, C., Martin, G. L. & Hanson, J. M. (2011). 'How robust are the findings of academically adrift?', *The Magazine of Higher Learning* 43: 20-4.
- Pike, G. R. (1995). 'The relationships between self reports of college experiences and achievement test scores.', *Research in higher education* 36: 1-22.