

# Resultater fra den første runden med referansemåling (benchmarking) i IMPI-prosjektet (mars 2011)

Rapport innenfor rammen av det europeiske  
prosjektet “**I**ndicators for **M**apping & **P**rofiling  
**I**nternationalisation” (IMPI)

## Generelle resultater

Det første forholdet vi kontrollerte, var hvor stor del av hver gruppe som leverte svarene, uavhengig av kvaliteten. I 4 av 5 grupper svarte alle medlemmer.

	Antall medlemmer <sup>1</sup>	Antall medlemmer som leverte data	%
Gruppe A	8	8	100%
Gruppe B	7	7	100%
Gruppe C	13 <sup>2</sup>	10	77%
Enkeltuniversitet 1	2	2	100%
Enkeltuniversitet 2	2	2	100%

Hver av gruppene eller enkeltuniversitetene valgte et individuelt sett indikatorer. Størrelsen på indikatorsettene varierte sterkt (mellom 18 og 56). Vanligvis ble det funnet data for hver indikator.

	Antall valgte indikatorer	Antall besvarte indikatorer	%
Gruppe A	28	28	100%
Gruppe B	18	18	100%
Gruppe C	32	32	100%
Enkeltuniversitet 1	28	28	100%
Enkeltuniversitet 2	56	53	95%

I mange tilfeller ble det kun gitt få svar på én enkelt indikator, slik at kun ett enkelt universitet kunne svare på alle indikatorer i alle tilfeller.

Gruppe A	88%
Gruppe B	82%
Gruppe C	88%
Enkeltuniversitet 1	100%
Enkeltuniversitet 2	90%

<sup>1</sup> Ved enkeltuniversitetet betyr "medlemmer" enheter eller avdelinger/fakulteter på universitetet/høyskolen  
<sup>2</sup> 7 av disse deltok ikke direkte, men via sin organisasjon ACUP.

## **Spesielle resultater**

### **Datakvalitet og metode**

Kvaliteten på avgitte data varierte sterkt. Ofte ble det i løpet av diskusjonen klart at institusjonene hadde oppfattet spørsmålene på forskjellige måter – til tross for at de hadde diskutert dem på forhånd. Noen ganger hadde alle forstått et ord, mens det andre ganger ble oppfattet helt ulikt. Et eksempel er begrepet ”internasjonal student”. Ved valg av indikator virket det klart for alle, men under sammenligningen av data viste det seg at begrepet i noen land ble oppfattet som en som kom fra utlandet, uavhengig av nasjonalitet, mens det i andre land var et spørsmål om vedkommenes pass.

Hovedresultatet var at i mange tilfeller kunne ikke opplysningene som ble samlet inn sammenlignes med hverandre: et typisk problem ved enkeltstående referansemålinger. Det ble klart for deltakerne at mer presisjon, mer diskusjon og eventuelt en ekstra innsamling av opplysninger kan være nødvendig for å gjøre resultatene helt sammenlignbare.

Metodisk sett ble det altså påvist hvordan dataanalysen avhenger av datakvaliteten, samt når visse instrumenter (som f.eks. gjennomsnitt) fungerer og når de ikke gjør det. Det ble også diskutert hvor grensene går for fordelene ved visualisering, i det minste i gruppen med brukbare data og de enkelte universitetene.

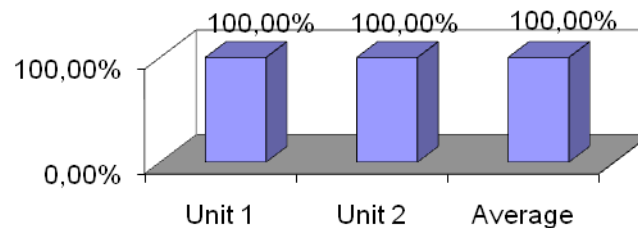
### **Forskjellene mellom egevaluering av enkelte universiteter og grupper**

Et viktig resultat av denne første testen er at enkeltstående referansemålinger fungerer langt bedre for egevaluering og rapportering internt i institusjonene enn de gjør for referansemålinger på tvers av institusjonene. Begge enkeltuniversitetene leverte verdifulle og tolkbare data. Verifisering av data ga inntrykk av å være enkel mellom avdelinger med sammenlignbar struktur (ett universitet brukte et matricesystem). For sentrale data var prosessen enda enklere. En ulempe var at fokus var rettet mot kun to avdelinger, men ettersom testen bare skulle være et eksempel og et forsøk, lærte universitetene nok til å utføre en test i større skala på

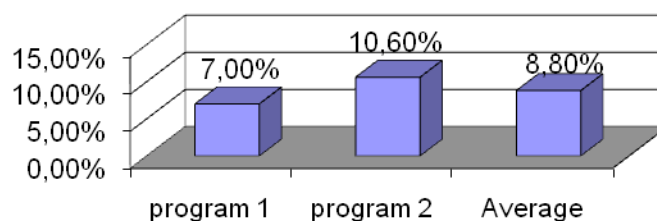
egenhånd. Det viste seg også at store indikatorsett (i ett tilfelle 56!) ikke er noe problem ved egnevalueringer.

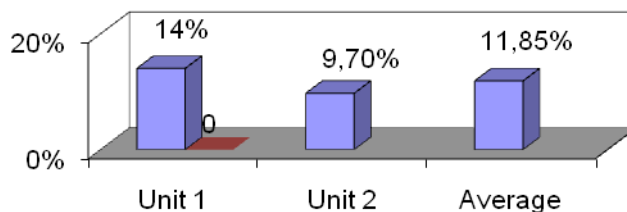
I tilfellet med ett enkelt universitet ga prosessen institusjonen muligheten til å sammenlikne aktiviteter og prestasjoner med henblikk på to programmer på bachelornivå innenfor et teknisk område. Det ble samlet inn data for alle de 28 indikatorene som ble undersøkt, noe som ga viktig innsikt i likheter og forskjeller når det gjelder graden og typen av internasjonalisering innenfor de to programmene. For eksempel har begge programmer en noe beskjeden andel innkommende utvekslingsstudenter sammenlignet med det totale antallet immatrikulerte studenter og utvekslingsstudenter som reiste ut, mens 100 % av studentene i begge programmene var påmeldt kurs i fremmedspråk (se tallene nedenfor).

**O-025 What is the proportion of students enrolled in your foreign language courses of all your students in this academic year?**



**O-075 What is the share of incoming exchange students of all students enrolled in your unit last year?**



**O-077 What is the share of outgoing students of all students enrolled in your unit in the last academic year?**

*O-025: Hvor stor andel av studentene var innskrevet på kurs i fremmedspråk blant alle studentene i dette akademiske året?*

*O-075: Hva er andelen innkommende utvekslingsstudenter blant alle innskrevne studenter i din enhet i fjor?*

*O-077: Hva er andelen utreisende studenter blant alle innskrevne studenter i din enhet i det siste akademiske året?*

For gruppene viste prosessen at det for en referansemålingsprosess også er svært viktig at alle medlemmer forstår begrensningene i den aktuelle prosessen og at forventningene er realistiske. En annen lærdom er at gruppe medlemmene må være mer oppmerksomme på sitt eget ansvar for å definere og klargjøre hva hver enkelt indikator tilsvarer i sin egen institusjon. Det krever mer tid enn man hadde i den første testrunden. I neste runde vil vi derfor sørge for at det er tid tilgjengelig for dette. Det anbefales også å bruke et mindre sett med indikatorer, slik at man får mer tid per indikator og mer intensive forberedelser ved hvert deltakende universitet før gruppearbeidet, spesielt for utvalget av indikatorer.

Vi hadde forventet at jo mer heterogen gruppen var, jo vanskeligere ville det bli å tolke dataene, ettersom forholdene ville være ulike og datakvaliteten derfor lavere. Testen viste imidlertid at dette ikke nødvendigvis er tilfelle. Den mest heterogene gruppen ga de nyttigste innsiktene for sine medlemmer, til tross for at det også oppsto misforståelser og at indikatorer måtte presiseres ytterligere.

Det var også påfallende at den gruppen som var vesentlig mindre, og dermed hadde et mer konsentrert indikatorsett (18), hadde en mer vellykket gjennomføring enn de

som valgte større sett (28-32). En av lærdommene fra testrunden er altså anbefalingen om å begrense gruppearbeidet til anslagsvis maksimalt 10 indikatorer.

En generell konklusjon for arbeidsgruppene er at det er bedre med en flertrinns referansemålingsprosess med en foreløpig egenevaluering av større indikatorsett, etterfulgt av et mer fokusert og bedre gjennomdiskutert indikatorsett.

En annen erfaring var at hvis et gruppemedlem ikke var til stede på det møtet der det endelige indikatorsettet ble bestemt, oppstod det problemer i etterkant, ettersom et slikt medlem da ville ha vanskeligheter med å oppfatte indikatoren i tråd med gruppen, vanskeligheter med å levere data og antagelig føle mindre tilknytning til disse indikatorene. En deltaker formulerte det slik: "Du må være sikker på at personen som har innhentet dataene har en egen interesse av å gjøre det."

### **Resultater fra indikatorene**

Den gruppen som genererte klart flest resultater fra referansemålingene, identifiserte mange viktige resultater angående indikatorene:

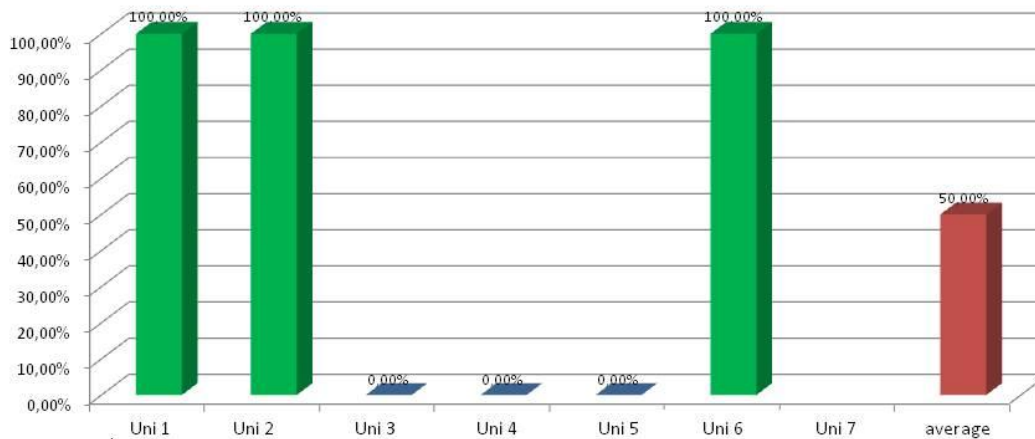
1. Formuleringen av en indikator er ikke klar nok: Ofte er årsaken bruken av spesialiserte ord som ikke er spesifisert i gruppen (f.eks. hva definerer vi som **totalbudsjett**?)
2. Indikatoren var klar, men det var ingen tilgjengelige data: Det vurderes derfor om det kan være til hjelp for institusjonen å vise at andre har relevante data og at deres egne prosedyrer derfor trenger forbedringer.
3. Alle leverte data og indikatoren var klar: En direkte sammenligning var derfor mulig, og ble utført.

I denne gruppens tilfelle, og ved det enkelte universitet, kunne vi identifisere former for god praksis, f.eks. ved spørsmål om hvordan man kan inkludere obligatoriske språkkurs i alle programmer, hvordan man kan rekruttere internasjonale medarbeidere eller hvordan man sikrer et mobilitetsvindu på alle grader. I enda flere tilfeller identifiserte vi god praksis i form av prosedyrer for å utlevere data som trengs for en slik øvelse som dette.

Noen eksempler viser dette:

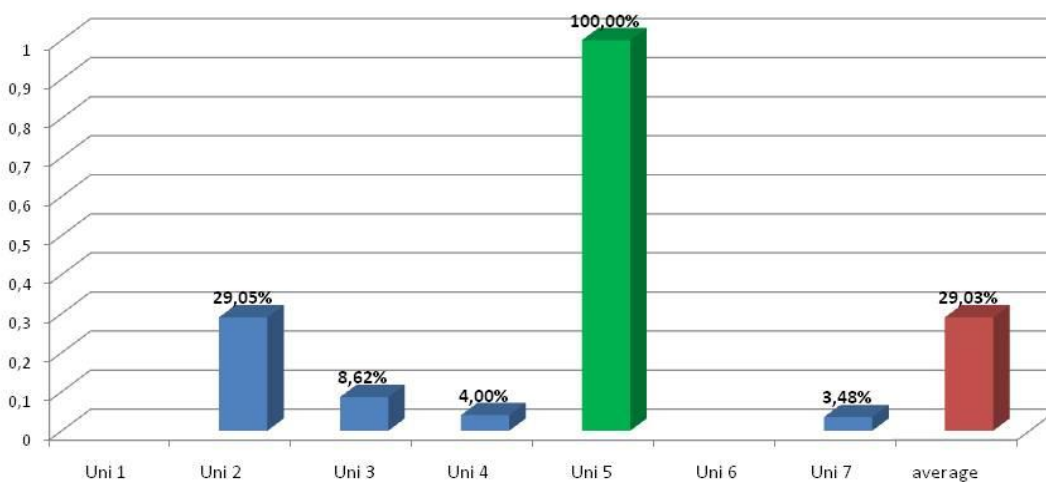
**Mobilitetsvindu som et spørsmål om "alt eller ingenting"**

**I-92: What is the proportion of curricula having mobility windows incorporated ?**



**Eksempel på fremragende prestasjoner**

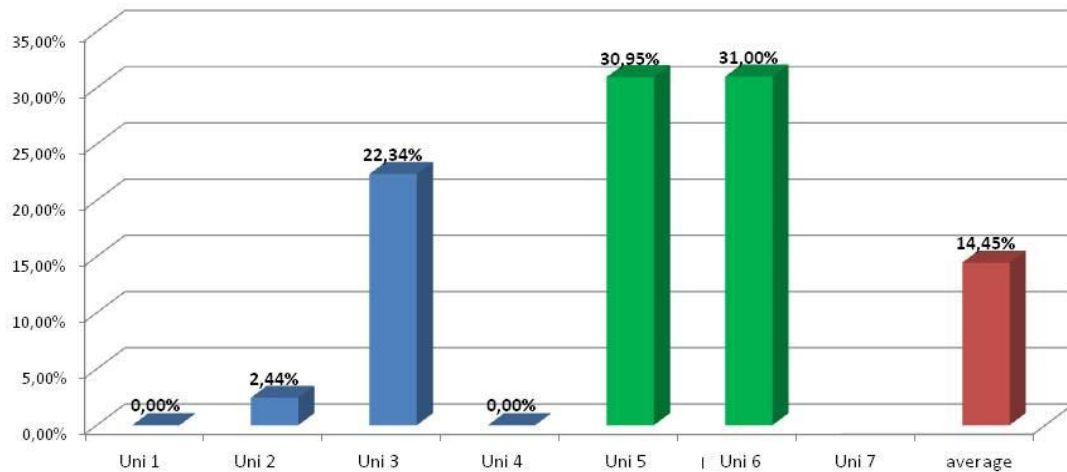
**O-25: What is the proportion of students enrolled in your foreign language courses of all your students?**





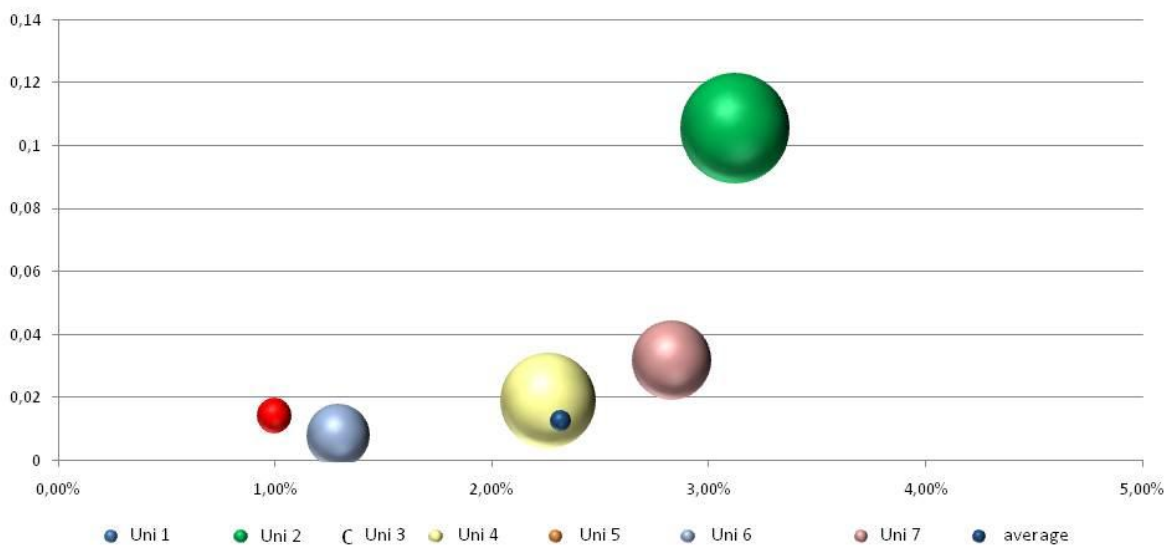
**Eksempel på et blandet resultat med to ledere og stor spredning**

**O-16: What is the proportion of Master's programmes taught in English if English is not the native language at the HEI (out of the total number of Master's programmes offered at your unit)?**



**Måter å presentere flerdimensjonale data**

**Relating new input (FTE share) as bubble size to two output indicators: x=incoming and y=outgoing**





*I-92: Hvor stor andel CV-er inneholder et element av mobilitet?*

*O-25: Hvor stor andel av studentene deltar på kurs i fremmedspråk?*

*O-16: Hvor stor andel av masterprogrammene undervises på engelsk hvis engelsk ikke er hovedspråket på universitetet (av det totale antallet masterprogrammer i enheten)?*

*"Mengden ny tilgang (FTE-andel) vist som boblestørrelse i forhold til to utgangsindikatorer  $x$ =innkommende og  $y$ =utreisende"*

## Verktøykasse

Verktøykassa kunne vise til ulike resultater. Testen viste at den fungerte for enkelte universiteter, selv i form av en grov skisse, for egnevaluering og til en viss grad også for gruppearbeidet, såfremt gruppen begrenset seg til færre indikatorer, var svært forsiktig med egne definisjoner og fant ulike måter å bruke mer tid på hver indikator.

Det ble dessuten klart at indikatorene må presiseres og strømlinjeformes ytterligere. De fleste grupper hadde forslag til forbedrede og mer presise formuleringer for å gjøre verktøykassa mer effektiv. Mer detaljerte definisjoner og ordforklaringer bidrar også til å redusere behovet for interne gruppediskusjoner. Samtidig ble det tydelig at verktøykassa ikke kan og ikke bør erstatte gruppens definisjoner, men må gjøres mer generell.

## Sammenfatning

Som forventet var den første testfasen svært nyttig for å utvikle verktøykassen ytterligere, og noe mindre nyttig på individuelt nivå for enkelte av partnerne. Det sistnevnte var både avhengig av deltakernes forventninger og av gruppestrukturen. Tatt i betraktning av at det handlet om et utviklingsprosjekt, produserte den første fasen alle resultater som trengtes. Dessuten valgte noen av deltakerne å bli med på andre testrunde og fordype sitt engasjement. Andre grupper bestemte seg for å kjøre noen data fra første runde på nytt etter å ha revidert indikatorer og oppfatninger. IMPI-teamet vil støtte dem med analysing av datamaterialet.