



LehrstellenPuls – Faktenblatt

Gewichtung und Repräsentativität der Ergebnisse

3. Dezember 2020

Forschungsteam

Thomas Bolli, Katherine M. Caves, Filippo Pusterla, Ladina Rageth, Ursula Renold, Aranya Sritharan, Sandra Trachsel Díaz-Tejeiro

Professur für Bildungssysteme, ETH Zürich

Forschungspartner

Urs Casty
Domenica Mauch

Yousty.ch, Professional.ch

Hintergrund

Um die Auswirkungen der COVID-19-Pandemie auf die berufliche Grundbildung in der Schweiz zu untersuchen, hat die Professur für Bildungssysteme der ETH Zürich in Kooperation mit Yousty.ch das Forschungsprojekt «LehrstellenPuls» initiiert. Der LehrstellenPuls erhebt seit April 2020 die monatliche Situation der Lehrbetriebe und Berufslernenden anhand einer Befragung von Lehrbetrieben. Dabei besteht die Stichprobe der monatlich kontaktierten Lehrbetriebe aus 44'660 E-Mail-Adressen, die von der Lehrstellenplattform Yousty.ch zu Verfügung gestellt werden. Bisher haben sich jeden Monat etwa 1'000 bis 2'600 Lehrbetriebe an der Befragung beteiligt.

Um zu sehen, wie zuverlässig und repräsentativ die Daten aus den LehrstellenPuls-Befragungen sind, müssen diese Daten mit der Grundgesamtheit – für welche Aussagen gemacht werden sollen – verglichen werden.¹ Da die Lehrbetriebe und Berufslernenden im Fokus des LehrstellenPulses stehen, stellen alle Lehrbetriebe in der Schweiz und ihre Berufslernenden die relevante Grundgesamtheit dar. Zudem gehen wir davon aus, dass sich die Auswirkungen der COVID-19-Pandemie auf die Lehrbetriebe und Berufslernenden je nach Betriebsgrösse, Region und Berufsfeld unterscheiden. Somit sind dies wichtige Vergleichskategorien, anhand welcher die Repräsentativität der Daten überprüft und gegebenenfalls korrigiert werden soll.

¹ Diese Faktenblatt ist eine Zusammenfassung des LehrstellenPuls-Berichts «Identifikation der Auswirkungen von COVID-19 auf die berufliche Grundbildung in der Schweiz. Bericht zu den Befragungen bei Lehrbetrieben von April bis August 2020 mit Fokus auf Gewichtungsmethodik und Heterogenität», welcher auf <https://lehrstellenpuls.ch/> heruntergeladen werden kann.

Datengrundlage

Abbildung 1 illustriert die Datengrundlage des LehrstellenPulses und deren Verhältnis zur Grundgesamtheit:

- Die graue Fläche stellt die **Grundgesamtheit** der Lehrbetriebe und ihrer Lernenden in der Schweiz dar.² Diese Angaben sind nach Betriebsgrößen, Kantonen und Berufen differenziert.
- Die pinke Fläche stellt die **Stichprobe** der kontaktierten Lehrbetriebe dar. Dabei gibt es verschiedene Gründe, weshalb sich diese von der Grundgesamtheit unterscheiden kann. So kann die Stichprobe zum Beispiel nicht alle Lehrbetriebe enthalten oder gewisse Betriebe sind aufgrund mehrerer Kontaktpersonen doppelt vertreten.
- Die **erweiterte Stichprobe** berücksichtigt, dass die Befragung nicht nur an die Stichprobe von Yousty geschickt, sondern auch auf weiteren Kanälen verbreitet wurde. Während bei der Stichprobe von Yousty Informationen zu den angebotenen Berufen, nicht aber zu weiteren Betriebsmerkmalen vorliegen, gibt es bei den durch die Weiterverbreitung erreichten Kontakten keine solchen Informationen.
- Die blaue Fläche stellt die **Beobachtungen** und damit alle Personen dar, die sich an den Lehrstellen-Puls-Befragungen beteiligt haben. Bei den

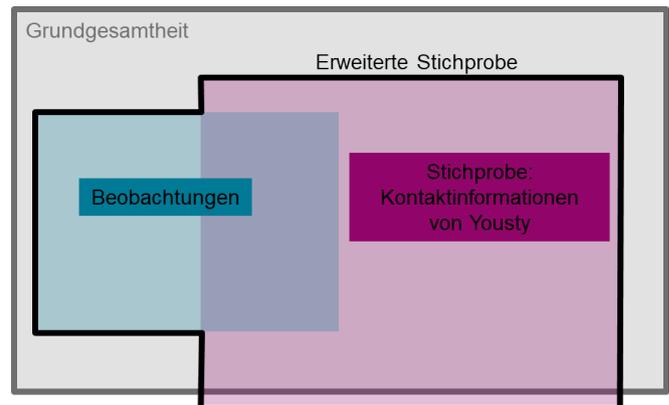


Abbildung 1 illustriert, das Verhältnis von Grundgesamtheit, Stichprobe und Beobachtungen, sie zeigt aber nicht die realen quantitativen Proportionen der verschiedenen Gruppen (eigene Darstellung).

Beobachtungen werden Informationen zu Betriebsgröße, Kanton und Beruf erhoben.

Damit sind die Beobachtungen des LehrstellenPulses eine Teilmenge der erweiterten Stichprobe und damit auch der Grundgesamtheit, für welche der LehrstellenPuls die Auswirkungen der Corona-Pandemie zeigen möchte – nämlich die Lehrbetriebe und Berufslernenden in der Schweiz. Um dennoch sicherzustellen, dass die Ergebnisse repräsentativ sind, können die Daten vor der Analyse gewichtet werden.

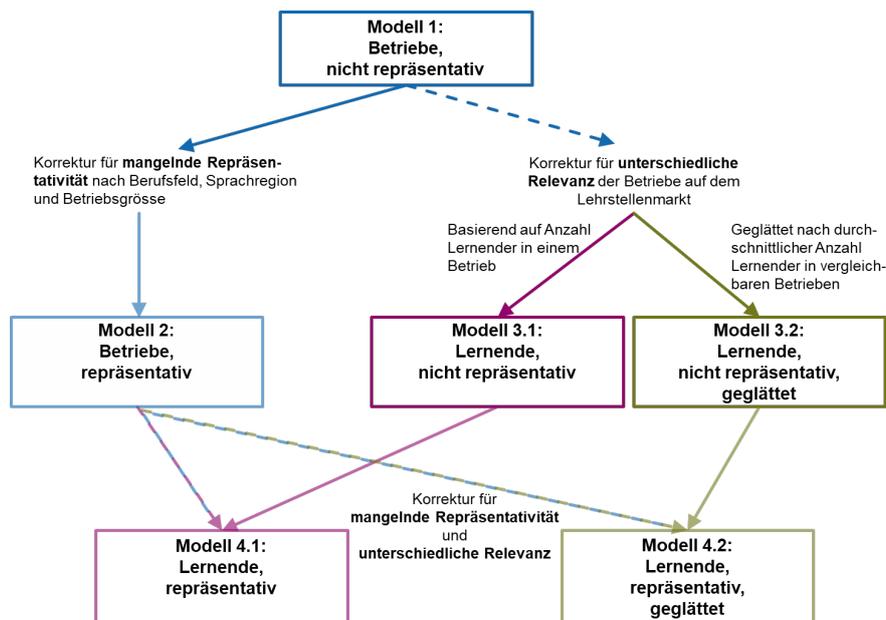


Abbildung 2 bietet eine Übersicht zu den möglichen Gewichtungsmodellen; Gewichte werden durch Berufsfeld, Sprachregion und Betriebsgröße definiert (eigene Darstellung).

² Die Verknüpfung der Statistik der beruflichen Grundbildung (SBG-SFPI) mit der Statistik der Betriebsstruktur (STATENT) liefert Angaben zur Anzahl aller Lehrbetriebe und Lernenden im Jahre 2017.

Beschreibung der Gewichtungsmodelle

Die Gewichtung der Daten verfolgt zwei Ziele:

Erstens soll die **Repräsentativität der Beobachtungen für die Grundgesamtheit** gewährleistet werden. Dies ist notwendig, weil sich die erweiterte Stichprobe von der Grundgesamtheit unterscheidet und weil sich das Antwortverhalten zwischen den Betrieben unterscheiden kann. Zweitens soll die **unterschiedliche Relevanz von Betrieben auf dem Lehrstellenmarkt** berücksichtigt werden. Wenn zum Beispiel ein Betrieb nur einen Berufslernenden ausbildet, ist dessen Einschätzung für den durchschnittlichen Lernenden weniger wichtig als die Einschätzung eines Betriebs mit 50 Lernenden.

Daraus können sechs mögliche Gewichtungsmodelle abgeleitet werden, die in Abbildung 2 dargestellt sind:

- Modell 1 **«Betriebe, nicht repräsentativ»**: Ausgangspunkt ist die Berechnung eines Durchschnitts der beobachteten Betriebe ohne Gewichtung.
- Modell 2 **«Betriebe, repräsentativ»**: Anhand der Korrektur für die mangelnde Repräsentativität in der Stichprobe, in der erweiterten Stichprobe und im Antwortverhalten berechnet man einen Durchschnitt der beobachteten Betriebe, der für die Grundgesamtheit repräsentativ ist.
- Modell 3 **«Korrektur für unterschiedliche Relevanz der Betriebe auf dem Lehrstellenmarkt»**
 - Modell 3.1 **«Lernende, nicht repräsentativ»**: Alternativ können Unterschiede in der Relevanz von Betrieben auf dem Lehrstellenmarkt berücksichtigt werden. Dazu werden Durch-

schnitte der beobachteten Lernenden berechnet, wobei jedoch die Repräsentativität nicht gewährleistet ist.

- Modell 3.2 **«Lernende, nicht repräsentativ, geglättet»**: Dieses Modell berechnet für vergleichbare Betriebe (z.B. grosse Betriebe in der Deutschschweiz) die durchschnittliche Anzahl Lernender in einem Betrieb und korrigiert anhand dieses Gewichts für die Relevanz des Betriebs.
- Modell 4 **«Korrektur für mangelnde Repräsentativität und unterschiedliche Relevanz»**
 - Modell 4.1 **«Lernende, repräsentativ»**: Anhand eines gewichteten Lernendendurchschnitts wird sowohl die mangelnde Repräsentativität der Beobachtungen als auch für die unterschiedliche Relevanz von Betrieben auf dem Lehrstellenmarkt korrigiert.
 - Modell 4.2 **«Lernende, repräsentativ, geglättet»**: Die Zusammenlegung der Gewichte für Repräsentativität und Relevanz vergleichbarer Betriebe ergibt einen gewichteten und geglätteten Lernendendurchschnitt.

Im ausführlichen Bericht werden die LehrstellenPuls-Ergebnisse zu ausgewählten Indikatoren anhand dieser sechs Modelle berechnet und die Ergebnisse verglichen. Diese Analysen liefern die Grundlage für die Auswahl des bei den zukünftigen LehrstellenPuls-Publikationen zu verwendenden Gewichtungsmodells.

Beispiel zur Analyse der Repräsentativität und Relevanz von Lehrbetrieben nach Berufsfeld

Illustrativ zeigt Abbildung 3 die Anteile der befragten Betriebe (Beobachtungen Betriebe, blau) und der Lernenden (Beobachtungen Lernende, hellrosa) im Vergleich zu den Lernenden in der Grundgesamtheit (dunkelrosa) für jedes Berufsfeld. Ein Vergleich der verschiedenfarbigen Balken zeigt, dass sich die Verteilungen nicht grundlegend unterscheiden.

Bei der Korrektur für die Repräsentativität kann die Gewichtung in zwei Richtungen gehen. Wenn der Anteil Lernender in den Beobachtungen grösser ist als deren Anteil in der Grundgesamtheit, muss dieser hinuntergewichtet werden (z.B. «Informatik» und «Elektrotechnik»). Wenn der Anteil Lernender in den Beobachtungen hingegen kleiner ist als in der Grundgesamtheit, wird dieser hochgewichtet (z.B. «Fahrzeuge» und «Verkauf, Einkauf»).

Auch die Korrektur der Relevanz kann in zwei Richtungen gehen: Wenn der Anteil der beobachteten Betriebe in einem Berufsfeld höher ist als der Anteil der beobachteten Lernenden, werden die Betriebe in diesem Berufsfeld hinuntergewichtet (z.B. bei «Wirtschaft, Verwaltung, Tourismus» und «Planung und Konstruktion»). Wenn der Anteil der Betriebe geringer ist als deren Anteil Lernender werden die Betriebe hochgewichtet (z.B. «Gesundheit» und «Elektrotechnik»).

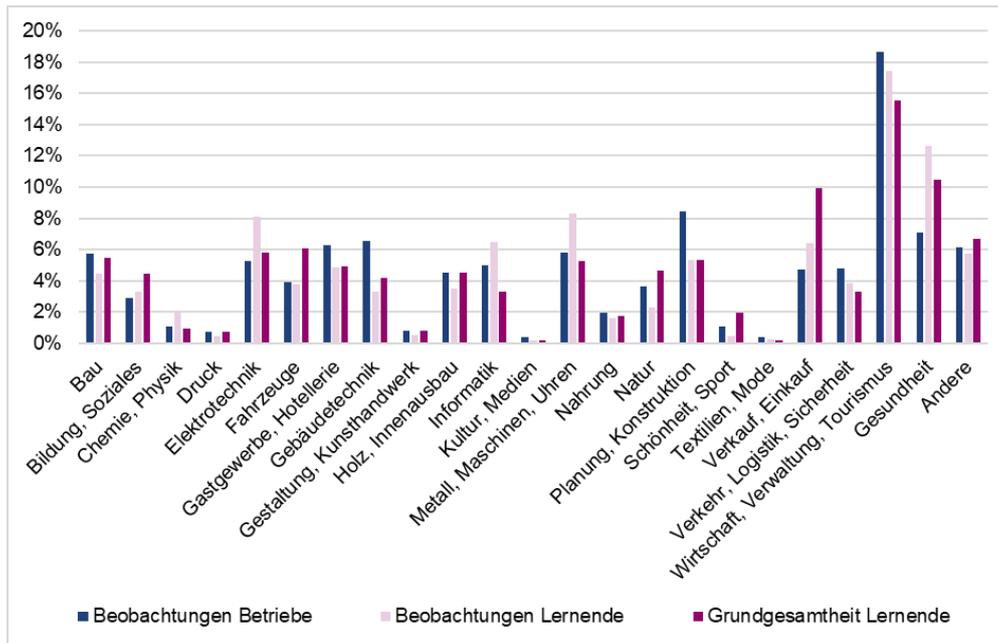


Abbildung 3 analysiert die Heterogenität der Repräsentativität und Relevanz anhand der Daten aus den LehrstellenPuls-Befragungen im April, Mai, Juni, Juli und August 2020. Sie zeigt pro Berufsfeld den Anteil der Betriebe in den Beobachtungen (blaue Säulen), den Anteil der Lernenden in den Beobachtungen (hellrosa Säulen) und den Anteil der Lernenden in der Grundgesamtheit (dunkelrosa Säulen); N= 10'135 beobachtete Betriebe / 90'778 beobachtete Lernende / 191'973 Lernende in der Grundgesamtheit Lernende.

Unterschiede in der Repräsentativität

Um zu sehen, ob die beobachteten Lernenden in Bezug auf Betriebsgrösse, Kanton und Berufsfeld repräsentativ sind, werden diese mit der Grundgesamtheit verglichen.³ Dabei zeigt sich, dass die Datengrundlage des LehrstellenPulses die Verteilungen der Lernenden nach Berufsfeld und nach Kanton insgesamt relativ gut abbildet. Das Berufsfeld «Verkauf, Einkauf» ist etwas unterrepräsentiert, während «Metall, Maschinen, Uhren» und «Informatik» etwas überrepräsentiert sind. Da mehrere Kantone der lateinischen Schweiz relativ wenige beobachtete Lernende aufweisen, ist diese Region insgesamt leicht unterrepräsentiert.

Allerdings zeigen sich grössere Unterschiede bei der Verteilung nach Betriebsgrösse. Kleinst- und Kleinbetriebe mit bis zu 50 Mitarbeitenden decken in den Beobachtungen deutlich weniger Lernende ab als in der Grundgesamtheit. Grossbetriebe beschäftigen hingegen mehr als die Hälfte der beobachteten Lernenden, während in der Grundgesamtheit nur rund ein Drittel der Lernenden in Grossbetrieben arbeitet.

Unterschiede in der Relevanz von Lehrbetrieben

Zur Analyse der Relevanz von Betrieben auf dem Lehrstellenmarkt werden die Unterschiede in den Verteilungen der beobachteten Betriebe und beobachteten Lernenden betrachtet, wobei diese bei Berufsfeld und Kanton relativ klein sind. Am stärksten sind die Unterschiede in den Berufsfeldern «Elektrotechnik», «Gebäudetechnik», «Metall, Maschinen, Uhren», «Gesundheit» und «Planung, Konstruktion» sowie in den Kantonen Basel-Stadt, Luzern, Waadt und Zürich.

Bei der Betriebsgrösse erhalten die Grossbetriebe durch die Berücksichtigung der Relevanz ein substanzial grösseres Gewicht. Während nur jede zehnte Beobachtung ein Grossbetrieb ist, decken diese über die Hälfte der beobachteten Lernenden ab.

³ Diese Analysen basieren auf aggregierten Daten der monatlichen LehrstellenPuls-Befragungen von April bis August 2020.

Gewichtung der LehrstellenPuls-Ergebnisse

Da die LehrstellenPuls-Daten die Verteilung von Lernenden nach Berufsfeld und Kanton relativ gut abbilden, sind die Ergebnisse ziemlich stabil, wenn eine Gewichtungsmethode zur Sicherstellung der Repräsentativität angewendet wird.

Hingegen führt die Korrektur für die Relevanz von Betrieben anhand der Anzahl Lernender im Betrieb teilweise zu grossen Unterschieden in den Ergebnissen. Dies ist insbesondere bei den berufsfeldspezifischen Analysen der Fall. Dabei stellt sich die Frage, ob diese Unterschiede entstehen, weil grössere Betriebe die Auswirkungen der Corona-Pandemie anders einschätzen oder weil die Stichprobengrösse von grösseren Betrieben zu klein ist, um mögliche Messfehler auszugleichen. Deshalb wird zusätzlich eine Gewichtungsmethode analysiert, welche die Anzahl Lernende in einem Betrieb glättet, indem die durchschnittliche Anzahl Lernender in vergleichbaren Betrieben (und nicht im jeweiligen Betrieb) verwendet wird. Diese Ge-

wichtungsmethode ergibt sehr ähnliche Resultate wie die Betriebsdurchschnitte, was darauf hindeutet, dass die Gewichtung anhand der Anzahl Lernender im Betrieb nicht die optimale Gewichtungsmethode darstellt.

Ausblick Berichterstattung

Vor diesem Hintergrund basieren die Ergebnisse zum LehrstellenPuls ab der November-Befragung hauptsächlich auf den repräsentativen und geglätteten Lernendendurchschnitten. Damit können wir für die mangelnde Repräsentativität korrigieren und die unterschiedliche Relevanz der Betriebe auf dem Lehrstellenmarkt berücksichtigen. Dabei weichen wir nicht allzu stark von den bisherigen Analysen ab, in welchen wir meistens nicht repräsentative Lernendendurchschnitte präsentierten (siehe Faktenblatt **«Forschungsdesign»** vom 7. Mai 2020). Für einzelne Indikatoren, bei welchen es inhaltlich mehr Sinn macht von Betrieben anstatt von Lernenden zu sprechen, werden die repräsentativen Betriebsdurchschnitte (bisher: nicht repräsentative Betriebsdurchschnitte) verwendet.

Gerne informieren wir Sie monatlich auf unserer Homepage über die neusten Ergebnisse:

www.lehrstellenpuls.ch →

Kontakt

ETH Zürich
Prof. Dr. Ursula Renold
Professur für Bildungssysteme
www.ces.ethz.ch →

LehrstellenPuls
info@lehrstellenpuls.ch →
www.lehrstellenpuls.ch →