



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

Erfassung möglicher Unterschiede von Respondern und
Non-Respondern in Leistungstestergebnissen

verfasst von / submitted by

Ulrike Goll

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Naturwissenschaften (Mag.rer.nat)

Wien / Vienna, 2015

Studienkennzahl /
degree programme code as it appears on the student record sheet: A 298

Studienrichtung /
degree programme as it appears on
the student record sheet: Diplomstudium Psychologie

Betreuer / Supervisor: Univ.-Prof. Dr. Mag. Klaus D. Kubinger

Vorwort

Ich möchte mich anfangs ganz herzlich bei meinem Diplomarbeitsbetreuer Herrn Prof. Dr. Kubinger für seine Unterstützung bedanken.

Danke auch an all jene Studienkollegen, die mir mit ihrem Rat zur Seite standen.

Vor allem gilt mein Dank meiner Familie, die meinen späten Wunsch, Psychologie zu studieren, akzeptierte, die Erfolge und auch Rückschläge während meines Studiums mit mir teilte und mir Rückhalt gab. Vielen Dank für alles!

Abstract

Often data of volunteers has to be used for the calibration of psychological diagnostic methods because it is the only data available. Under the assumption that so-called “responders” differ from “non-responders” in their intelligence performance the application of a method calibrated with such data to the whole population seems problematic. Moreover, the generalization of study results based on data of voluntary participants has to be viewed critically if non-responders and responders have different characteristics that are relevant for the study. In this study possible differences with regard to cognitive skills were examined. First, 105 students of Lower Austrian secondary schools had to do three performance tests. In addition, the “performance-based self-worth” was explored to control for possible influences of this variable. Furthermore, the willingness to participate in further surveys was requested. Second, differences between responders and non-responders were analysed: participants who actually answered to another questionnaire that was sent by e-mail were classified as responders. The group of non-responders included people that did not answer to a further questionnaire. No significant difference of performance in the intelligence test was found between responders and non-responders. A possible influence of the covariate “performance-based self-esteem” has not been confirmed. A generalization of data based on studies with volunteers therefore seems unproblematic.

Abstract

Für die Eichung psychologisch-diagnostischer Verfahren stehen meist nur Daten freiwilliger Teilnehmer zur Verfügung. Unter der Annahme, dass sich diese sogenannten Responder in ihrer Intelligenzleistung von Non-Respondern (Teilnahme-Verweigerern) unterscheiden, ist die Anwendung auf solche Art geeichter Verfahren in der Gesamtbevölkerung problematisch. Auch die Verallgemeinerung von Studienergebnissen, die nur auf Daten freiwilliger Teilnehmer basieren, ist kritisch zu betrachten, wenn sich Non-Responder in für die Forschungsfrage relevanten Eigenschaften von Respondern unterscheiden. In dieser Arbeit werden mögliche Unterschiede von Respondern und Non-Respondern hinsichtlich kognitiver Fähigkeiten untersucht. Zunächst wurden 55 Maturantinnen und 50 Maturanten aus fünf niederösterreichischen Schulklassen je drei Leistungstests vorgegeben. Zusätzlich wurde der leistungsbezogene Selbstwert erfasst, um mögliche Einflüsse desselben auf die Ergebnisse kontrollieren zu können. Auch das Ausmaß der Bereitschaft, an weiteren Befragungen teilzunehmen, wurde erfragt. Danach erfolgte die Unterscheidung von Respondern und Non-Respondern: Teilnehmer, die tatsächlich einen weiteren, zugesandten Fragebogen bearbeiteten, wurden in die Gruppe der Responder eingeordnet. Zu den Non-Respondern wurden alle gezählt, die diesen nicht beantworteten. Es konnten keine signifikanten Unterschiede im Leistungsniveau der beiden Gruppen gefunden werden. Auch ein möglicher Einfluss der Kovariate leistungsbezogenes Selbstwertgefühl konnte nicht bestätigt werden. Die Erfassung der Bereitschaft, an weiteren Befragungen teilzunehmen, zeigte, dass die Non-Responder dieser Studie mehrheitlich passive Non-Responder waren.

Eine Generalisierung von Daten basierend auf Untersuchungen mit Freiwilligen scheint somit unproblematisch.

Inhaltsverzeichnis

I. Einleitung.....	Seite 8
II. Theoretischer Teil.....	Seite 10
1. Überblick über bereits bekannte Divergenzen.....	Seite 10
2. Kognitive Leistung von Respondern und Non-Respondern.....	Seite 11
3. Einfluss des Selbstkonzepts.....	Seite 13
4. Grad der Freiwilligkeit.....	Seite 14
III. Empirischer Teil.....	Seite 14
5. Ziel der Untersuchung (Hypothesen).....	Seite 14
6. Methode.....	Seite 15
6.1. Untersuchungsplan.....	Seite 15
6.2. Erhebungsinstrumente.....	Seite 16
6.2.1. Modul 4 des Wilde-Intelligenz-Test 2	Seite 16
6..2.2. Multidimensionale Selbstwertkala: Leistungsbezogene Selbstwertschätzung.....	Seite 18
6.2.3. Erfassung des Grades der Freiwilligkeit.....	Seite 19
6.2.4. AKZEPT! Fragebogen.....	Seite 20
6.3. Durchführung der Untersuchung.....	Seite 21
6.4. Stichprobe.....	Seite 23
6.4.1. Stichprobenumfang.....	Seite 23
6.4.2. Stichprobenzusammensetzung.....	Seite 24
7. Ergebnisse.....	Seite 25
7.1. Auswertung.....	Seite 25
7.2. Ergebnisse: durch Unterscheidung von Respondern und Non-Respondern ermittelter Effekt.....	Seite 26
7.3. Ergebnisse: einfache multivariate Kovarianzanalyse.....	Seite 26
7.4. Ergebnisse zum angegebenen Grad der Freiwilligkeit.....	Seite 27
7.4.1. Ergebnisse zum Grad der Freiwilligkeit der Non-Responder...	Seite 28
7.4.2. Ergebnisse zum Grad der Freiwilligkeit der Responder.....	Seite 28
8. Diskussion und Ausblick.....	Seite 29
9. Zusammenfassung.....	Seite 31
10. Literaturverzeichnis.....	Seite 33
Anhang 1.....	Seite 37

Anhang 2.....	Seite 38
Anhang 3.....	Seite 39
Anhang 4.....	Seite 40
Lebenslauf.....	Seite 41

I. Einleitung

Gut gestaltete Befragungen erweisen sich als effektive Forschungsinstrumente. Wenn die erhobenen Daten aus einer sorgfältig gewählten, repräsentativen Stichprobe stammen, können sie auf die gesamte, interessierende Population verallgemeinert werden. Personen, die zur Befragung vorgesehen sind, bilden die sogenannte Ausgangsstichprobe. Der Anteil tatsächlich interviewter Personen an der Ausgangsstichprobe – die sogenannte Ausschöpfungsquote – erreicht bei sozial-wissenschaftlichen Umfragen fast nie hundert Prozent. In jeder Umfrage gibt es Personen, die entweder nicht erreicht werden können oder solche, die die Mitarbeit versagen. Die Ausschöpfungsquote schwankt dabei beträchtlich und reicht von ungefähr 70% bei mündlichen Befragungen bis zu lediglich 5-10% bei den meisten schriftlichen Befragungen (Kurz, 1987). Generell werden für Länder, die über langjährige Traditionen und Erfahrungen im Bereich der Umfrageforschung verfügen, seit den 1950er Jahren tendenziell steigende Ausfälle, also eine Zunahme von Non-Response bei Bevölkerungsumfragen, festgestellt (Neller, 2005).

Abgesehen von der Vergrößerung des statistischen Zufallsfehlers, hätte dies keinerlei Beeinträchtigung der Aussagekraft der Umfrageergebnisse zur Folge, wenn sich die Antwortreaktionen der nichterreichten Personen nicht vom Antwortverhalten der Erreichten unterscheiden würden (Kurz, 1987). Schumann (2012) unterscheidet dabei unsystematische und systematische Ausfälle. Während die Gründe für unsystematische Ausfälle in keinem Zusammenhang mit der Forschungsfrage stehen, hängen bei systematischen Ausfällen die Ursachen der Nicht-Teilnahme mit den untersuchten Merkmalen und ihren Ausprägungen zusammen. Ob und in welchem Maße die Ausfälle bei einer Stichprobe systematischer oder unsystematischer Natur sind, lässt sich in den seltensten Fällen empirisch prüfen, da hierzu die Daten der nicht teilnehmenden Personen nötig wären. Häufig hängen die Ursachen für die Nichterreichbarkeit bzw. Bereitschaft zur Mitarbeit mit den untersuchten Einstellungs- und Verhaltensmustern zusammen. So füllen nach Kurz (1987) die Teilnehmer an einer schriftlichen Befragung den Fragebogen deshalb aus, weil sie aus irgendwelchen Gründen dem Untersuchungsthema ein erhöhtes Interesse entgegenbringen. In welche Richtung

eine Verzerrung der Ergebnisse durch systematische Ausfälle – der sogenannte Non-Response-Bias - wirkt, kann meist nur vermutet werden.

Bei der Normierung psychologisch-diagnostischer Verfahren ist man größtenteils auf die Mithilfe Freiwilliger angewiesen. Es kann aber zu einer fehlerhaften Interpretation der Ergebnisse kommen, wenn Verfahren, die mit Hilfe von Respondern geeicht wurden, zur Testung oder Befragung von Non-Respondern eingesetzt werden. Die Einschätzung des tatsächlichen Ausmaßes einer solchen Verzerrung, setzt Wissen über systematische Unterschiede von Respondern und Non-Respondern voraus. In einer Arbeit von Tiapal (2011) wurde versucht, auch Daten von Teilnahmeverweigerern zu erhalten, indem Schulklassen einem Persönlichkeits- und Motivationstest unterzogen wurden und die Nicht-Teilnahme an einem weiteren Projekt als Zuordnungskriterium zur Gruppe der Non-Responder diente. Tiapal konnte nur bei der Persönlichkeit, nicht aber bei den Motiven Unterschiede zwischen den Gruppen finden.

In der vorliegenden Arbeit sollen mögliche Unterschiede von Respondern und Non-Respondern in Leistungstestergebnissen untersucht werden. Im ersten Schritt sollen kognitive Fähigkeiten aller Teilnehmer erfasst werden. Erst danach erfolgt die Unterscheidung in Teilnahmeverweigerer und Freiwillige. Mögliche Unterschiede der beiden Gruppen in ihren kognitiven Fähigkeiten werden geprüft.

Im theoretischen Teil dieser Arbeit erfolgt eine kurze Erläuterung zu den interessierenden Merkmalen. Im empirischen Teil werden das Ziel dieser Untersuchung und die dabei angewandte Methode näher erläutert. Die verwendeten psychologisch-diagnostischen Verfahren und auch die Stichprobe werden ausführlich beschrieben. Danach erfolgt die Beschreibung der statistischen Analysen und der Ergebnisse, die abschließend diskutiert werden.

II. Theoretischer Teil

Die Gründe für das Ablehnen einer Teilnahme an Befragungen können sehr unterschiedlich sein. Neller (2005) nennt als Ursachen für die Zunahme der Non-Responder bei Bevölkerungsumfragen unter anderem allgemeine gesellschaftliche Entwicklungen, wie abnehmende Partizipationsbereitschaft, zunehmende Individualisierung, Stress und Zeitmangel, verändertes Freizeitverhalten und erhöhte Mobilität. Als weitere Gründe für Teilnahmeverweigerung sieht er die zunehmende Debatte um Datenschutz und das steigende Bedürfnis danach sowie eine gewisse Ermüdung der Bürger durch die stetig steigende Zahl von Studien im Bereich der Marktforschung.

Es stellt sich die Frage nach den Hintergründen, Prozessen und Mechanismen, die die individuelle Entscheidung jedes einzelnen, teilzunehmen oder zu verweigern, prägen. Im Folgenden werden bisherige Erkenntnisse zu diesem Thema dargelegt und die Bedeutung der dieser Studie zugrunde liegenden Konstrukte für die Non-Response-Problematik verdeutlicht.

1. Überblick über bereits bekannte Divergenzen

Verschiedene Studien zum Thema Non-Response beschäftigen sich mit Unterschieden von Respondern und Non-Respondern hinsichtlich soziodemographischer Daten, unterschiedlicher Persönlichkeitsmerkmale und Motive. Nach Marcus und Schütz (2005) zeigten Responder höhere Werte in Extraversion, Offenheit für neue Erfahrung und Narzissmus. Carlo, Okun, Knight und Guzmann (2005) fanden ähnliche Ergebnisse, wobei sich zwischen Verträglichkeit und prosozialer Wertmotivation ein starker Zusammenhang zeigte. Beide sind signifikante Prädiktoren für freiwilliges Verhalten. Auch Lonqvist et al. (2007) fanden höhere Werte in Gewissenhaftigkeit, Extraversion und Verträglichkeit und niedrigere Werte für Neurotizismus bei Freiwilligen. Sie weisen darauf hin, dass durch die Verwendung Freiwilliger für die Eichung psychologisch-diagnostischer Verfahren die Ergebnisse der Normierung verzerrt sein können.

Lyness und Brumit (2007) fanden, dass Frauen eher zu Respondern zählen und die Freiwilligenrate in Ländern mit höherer Gleichberechtigung größer ist. Darüber hinaus zeigten Responder einen höheren Bildungsgrad und einen höheren sozio-ökonomischen Status. Der typische Freiwillige ist an dem Umfragethema interessiert und glaubt, dass er durch seine Teilnahme zu Veränderungen beiträgt. Auch nach Burkell (2003) glauben Responder stärker, dass die Studienergebnisse etwas bewirken können. Unter den Respondern finden sich verstärkt Personen mit hohem sozialem Engagement (Porter & Whitcomb, 2005). Befragte mit einer positiven und kognitiv zugänglichen Haltung gegenüber Umfragen zeigen eine kooperative Einstellung (Stockè, 2006).

Empirische Untersuchungen zum besseren Verständnis der Entscheidungen, die Handlungen vorausgehen, haben zudem gezeigt, dass individuelles Verhalten in hohem Maße von der Wahrnehmung des sozialen Umfeldes beeinflusst wird (Aronson, Wilson & Akert, 2008). Nicht nur objektive Bedingungen bestimmen demnach die Entscheidung teilzunehmen, sondern auch die subjektive Wahrnehmung der Umgebung spielt eine Rolle. Kognitive Fähigkeiten wie Wahrnehmung, Aufmerksamkeit, aber auch Denken, Erinnern ermöglichen Menschen, Situationen und Herausforderungen, wie z.B. die Bitte um Teilnahme an einer Studie, einzuordnen und zu bewerten. Das führt zu der Frage, ob Unterschiede in den kognitiven Fähigkeiten für die Frage der Freiwilligkeit von Bedeutung sind.

2. Kognitive Leistung von Respondern und Non-Respondern

Intelligenz, das Schlüsselmerkmal für Erfolg in Alltag und Beruf, ist die Fähigkeit zur Anpassung an neuartige Bedingungen und die Fähigkeit zur Lösung neuartiger Probleme (Kubinger & Jäger, 2003). Ein wesentliches Kriterium der Intelligenz ist die Geschwindigkeit der Informationsverarbeitung. Je schneller diese erfolgt, umso mehr Information kann aufgenommen werden. Summiert über einen langen Zeitraum hinweg, sollen dadurch erhebliche interindividuelle Differenzen im Wissen und in den Fähigkeiten resultieren. Dementsprechend finden sich in Leistungstests Aufgaben mit einem festgelegten Zeitlimit.

Nach Thurstons Modell der „Primary Mental Abilities“ setzt sich Intelligenz aus sieben spezifischen Fähigkeiten zusammen. Räumliches, sprachliches und schlussfolgerndes Denken sind drei davon. Induktives Schlussfolgern bedeutet etwa, dass jemand, der die Bearbeitung eines Fragebogens positiv wahrnimmt, eher bereit ist, an weiteren Befragungen teilzunehmen. Umgekehrt kann eine Person, die bereits mehrfach schlechte Erfahrungen mit Testungen oder Befragungen gemacht hat, eine weitere Teilnahme verweigern, ohne sich näher mit der Anfrage auseinanderzusetzen, weil sie von ihren Erfahrungen auf den Einzelfall schließt (deduktives Schlussfolgern).

In modernen Gesellschaften müssen aufgrund zahlreicher Verhaltensanforderungen viele Entscheidungen getroffen werden. Dies führt zu automatischen Einwilligungs- bzw. Verweigerungsverhalten mithilfe heuristischer Regeln. Dadurch können aufwändige kognitive Prozesse wie das Abwägen von Vorteilen und Nachteilen einer Teilnahme vermieden werden. Zu diesen heuristischen Regeln gehört etwa das Prinzip der Reziprozität. Reziprozität bedeutet, dass auf positives Verhalten mit positivem Verhalten reagiert wird. Durch optimale Gestaltung einer Befragung kann demnach die Teilnahmebereitschaft gesteigert werden. Haunberger (2011) weist darauf hin, dass die aufwendige Gegenüberstellung von Aufwand und Nutzen einer Teilnahme durch Befragte oftmals entfällt, wenn sie z.B. mit dem Befragungsthema vertraut sind. Sie entscheiden sich möglicherweise automatisch-spontan für die Teilnahme an einer Befragung oder Testung. Mit bewussten Entscheidungsprozessen ist zu rechnen, wenn die Handlungssituation als wichtig angesehen wird und die Konsequenzen kalkulierbar erscheinen. Welche Strategie zur Anwendung kommt, hängt unter anderem von den kognitiven Fähigkeiten ab (Schnell, 1997).

Neller (2005) berichtet, dass vor allem höher Gebildete für die Teilnahme an Befragungen gewonnen werden können, insbesondere wenn die Erhebungen inhaltlich anspruchsvoll sind. Bei einer Untersuchung von Schräpler (2000) findet sich die Vermutung, dass ein höheres Bildungsniveau die Erfolgsaussichten für die Motivation zur Interviewteilnahme bei Frauen erhöht. Auch nach Schützenmeister (2002) sind höher Gebildete häufiger bereit, sich erneut befragen zu lassen. Sie können eher den Sinn von Befragungen zu Sachverhalten sehen, die über ihre unmittelbare Lebenswelt hinausgehen. Demgegenüber fand Gottschling (2013) in

einem Quasiexperiment zur Trainierbarkeit von Textverständnis hinsichtlich Geschlecht, Schulleistungen und Sozialstatus keine Unterschiede zwischen Freiwilligen und Nichtfreiwilligen.

Im Rahmen einer Diplomarbeitsstudie am Arbeitsbereich Psychologische Diagnostik der Fakultät für Psychologie der Universität Wien konnte Unterhofer (2009) keinen signifikanten Einfluss des Grades der Freiwilligkeit auf die Testergebnisse aus drei Leistungstests finden. Die Teilnehmer wurden dabei nach dem Zeitpunkt der Teilnahme nach mehreren Aufrufen in verschiedene Gruppen unterteilt. Es wurden aber keine absoluten Non-Responder erfasst.

3. Einfluss des Selbstkonzepts

Das Selbstkonzept einer Person ist ihre Vorstellung über die eigenen Fähigkeiten (Benölken, 2014). Es wird maßgeblich aus Rückmeldungen von Bezugspersonen und damit verbundenen Attributionen geprägt und enthält kognitive und affektive Anteile. Der Erfolg einer Person bei der Lösung einer Aufgabe kann dementsprechend darauf zurückgeführt werden, dass sie ausreichend begabt ist oder darauf, dass sie sich sehr angestrengt hat. Wenn intelligente und erfolgreiche Personen ihre Erfolge anzweifeln, weil sie ihre Leistungen nicht ihren intellektuellen Fähigkeiten zuschreiben, bezeichnet man das als Impostor Phänomen. Die Bedeutung dieser Ursachenzuschreibung für Leistungserfolg zeigt sich an den Ergebnissen einer Studie von Wiesner (2011). Er nennt das Impostor Phänomen als Erklärungsansatz für die Unterrepräsentanz von Frauen in akademischen Spitzenpositionen. Trotzdem also grundsätzlich wissenschaftlicher Konsens über gleiche Begabungspotenziale in allen akademischen Bereichen bei den Geschlechtern besteht, finden sich letztendlich weniger Mädchen und Frauen in mathematischen Ausbildungsgängen und Berufen (Benölken, 2014).

Die Überzeugung in Beruf und Arbeit leistungsfähig zu sein, wirkt sich demnach vermutlich auch auf den Erfolg bei Leistungstests aus und soll daher als Kovariate in dieser Arbeit miteinbezogen werden.

4.. Grad der Freiwilligkeit

Nach Rogelberg et al. (2003) können Personen, die nicht freiwillig an Befragungen oder Testungen teilnehmen, in "aktive" und "passive" Non-Responder unterteilt werden. Personen, die generell eine Teilnahme ablehnen, können demnach als „aktive“ Non-Responder bezeichnet werden. Personen, die grundsätzlich einer Teilnahme zustimmen, später aber doch nicht teilnehmen, werden als „passive“ Non-Responder erfasst. Während sie aufgrund äußerer Umstände nicht mehr teilnehmen, treffen aktive Non-Responder eine bewusste Entscheidung (Rogelberg et al., 2003). Diese Unterscheidung ist von Bedeutung, weil es schwer ist, Daten von Teilnahmeverweigerern zu bekommen. Da passive Non-Responder in Selbstbeschreibungen zur Persönlichkeit stärker denen von Freiwilligen ähneln (Marcus & Schütz, 2005), werden ihre Daten zur Klärung der Non-Response-Problematik miteinbezogen. Daten von späteren Non-Respondern erhält man bei der Frage nach der erklärten Bereitschaft zu einer erneuten Teilnahme in der Ersterhebung (Schützenmeister, 2002). Die Bereitschaft zur weiteren Teilnahme soll in dieser Arbeit mit erfasst werden.

III. Empirischer Teil

5. Ziel der Untersuchung (Hypothese/n)

Das Ziel dieser empirischen Untersuchung ist die Prüfung der Frage, ob sich Responder in ihren Leistungstestergebnissen signifikant von Non-Respondern unterscheiden.

Im ersten Teil der Untersuchung wird die schlussfolgernde, sprachliche und räumliche Fähigkeit von allen Teilnehmern mit Hilfe eines geeigneten psychologischen-diagnostischen Verfahrens erhoben. Zusätzlich zu den tatsächlichen Leistungen wird die leistungsbezogene Selbstwertschätzung erhoben. Es handelt sich dabei um eine variable Größe, die die Ergebnisse der Leistungstests verzerren kann. Dies soll bei der Auswertung der Ergebnisse berücksichtigt werden. Schließlich wird auch die Bereitschaft, an weiteren

Untersuchungen teilzunehmen, erhoben, um einen möglichen Zusammenhang zwischen dem Grad der Freiwilligkeit und kognitiver Leistung zu untersuchen.

Im zweiten Teil der Untersuchung erfolgt die Unterscheidung von Respondern und Non-Respondern. Alle Teilnehmer an den Leistungstests erhalten mit der Zusendung der Ergebnisse einen weiteren Fragebogen. Teilnehmer, die der Bitte nachkommen, diese zweite Befragung zu bearbeiten, werden der Gruppe der Responder zugerechnet. Ihre Testergebnisse werden mit den Ergebnissen der Teilnehmer, die den zweiten Fragebogen nicht beantwortet haben, den Non-Respondern, verglichen.

Daraus ergeben sich die folgenden Hypothesen:

H0: Unter Kontrolle der Kovariate Leistungsbezogener Selbstwert gibt es keine Unterschiede zwischen den Testergebnissen der Responder und der Non-Responder.

H1: Unter Kontrolle der Kovariate Leistungsbezogener Selbstwert gibt es Unterschiede zwischen den Ergebnissen der Responder und der Non-Responder.

6. Methode

Es folgt eine Beschreibung des Vorgehens, der dabei verwendeten psychologisch-diagnostischen Verfahren, der Durchführung sowie der Stichprobe.

6.1. Untersuchungsplan

Der Einfluss des Faktors Responder/Non-Responder auf die Ergebnisse aus drei Leistungstests soll untersucht werden. Der Einfluss der Kovariable Leistungsbezogener Selbstwert ist auszuschalten: Unterschiede hinsichtlich der

Testwerte sollen um den vom leistungsbezogenen Selbstwert verursachten Beitrag bereinigt werden. Die Bereitschaft zu weiteren Teilnahmen - der Grad der Freiwilligkeit - wird zusätzlich erfasst.

6.2. Erhebungsinstrumente

Als Erhebungsinstrument dient ein Leistungstest zur Erfassung der Fähigkeit im schlussfolgernden, sprachlichen und räumlichen Denken mit limitierter Bearbeitungszeit. Die Überzeugung der Teilnehmer leistungsfähig zu sein, wird mit einem Selbstbeschreibungsfragebogen erfasst. Zusätzlich wird die Bereitschaft, an weiteren Befragungen teilzunehmen, mit einer sieben-stufigen Skala erfasst.

Zur Unterscheidung von Respondern und Non-Respondern erhalten alle Teilnehmer der Testung mit den Ergebnissen einen weiteren Fragebogen, den AKZEPT! Fragebogen, zugesandt. Dessen Ergebnisse sind für die Studie nicht von Bedeutung. Von Bedeutung ist lediglich, ob er bearbeitet und rückgesandt wird.

Alle Verfahren werden anschließend detailliert beschrieben.

6.2.1. Modul 4 des Wilde-Intelligenz-Test 2

Der WIT-2 basiert auf einer Erweiterung der Primary Mental Abilities, dem sogenannten modifizierten Modell der Primary Mental Abilities und aktuellen Erkenntnissen der Intelligenzforschung.

Für die aktuelle Fragestellung von Bedeutung ist, dass das Verfahren kognitive Fähigkeiten differenziert misst. Dabei ermöglicht der WIT-2 eine flexible und zeitökonomische Testgestaltung durch ein Baukastensystem mit acht Modulen. Das für die Studie herangezogene Modul 4 kann in einer Unterrichtseinheit ohne Zeitdruck durchgeführt werden. Der überwiegende Teil der Normdaten für den WIT-2 wurde im Kontext von beruflichen Bewerbungssituationen erhoben und nicht anhand anonymer Gruppen von Schülern oder Studierenden. Man kann daher von Normierungswerten ohne Non-Responder-Bias ausgehen. Der WIT-2 erfreut sich bei Testteilnehmern nachweislich einer hohen Akzeptanz (Kersting, Althoff & Jaeger, 2008).

Modul 4 erfasst schlussfolgerndes Denken anhand von:

- Analogien: sprachliches und schlussfolgerndes Denken werden anhand von 20 Items in einer Bearbeitungszeit von 4,30 Minuten erfasst. Auf der linken Seite eines Gleichheitszeichens sind zwei Wörter vorgegeben, die in einer bestimmten Beziehung zueinander stehen, auf der rechten Seite ein Wort. Von fünf Wahlwörtern ist dasjenige auszuwählen, das auf der rechten Seite des Gleichheitszeichens eine analoge Beziehung herstellt.
- Abwicklungen: räumliches und schlussfolgerndes Denken werden anhand von 20 Items in einer Bearbeitungszeit von 9 Minuten erfasst. Zu einer aus mehreren Flächen mit verschiedenen Zeichen bestehenden Faltvorlage ist aus fünf vorgegebenen Körpern derjenige herauszufinden, der sich aus der Faltvorlage herstellen lässt.
- Zahlenreihen: schlussfolgerndes Denken wird anhand von 20 Items in einer Bearbeitungszeit von 10 Minuten erfasst. Vorgegeben ist eine Folge von Zahlen, die nach einer Regel aufgebaut ist; auf dem Antwortbogen sind die Ziffern der Zahl durchzustreichen, die als nächstes Glied der Reihe folgen müsste.

Das Verfahren erfüllt folgende Testgütekriterien:

Objektivität: Die Durchführungsobjektivität wird durch das Manual und das zusätzlich vorliegende Instruktionssheft für den Testleiter sicherzustellen versucht. Die Auswertungsobjektivität wird anhand vorliegender Schablonen und übersichtlicher Ergebnisbögen gewährleistet (Kersting et al., 2008, S. 43).

Reliabilität: Die interne Konsistenz liegt für Analogien bei 0.79 nach Cronbachs alpha, die Retest-Reliabilität nach 9 Monaten beträgt $r_{tt} = 0.75$. Bei Abwicklungen ergibt sich ein Cronbachs Alpha von 0.84; $r_{tt} = 0.84$. Für Zahlenreihen liegt Cronbachs Alpha bei 0.87 und $r_{tt} = 0.72$ (Kersting et al., 2008, S. 43).

Validität: Durch die Bestimmung der Zusammenhänge zwischen Aufgaben des WIT-2 einerseits und Aufgaben aus 17 anderen Testverfahren andererseits wurde die konvergente und diskriminante Validität empirisch ermittelt (Kersting et al., 2008, S. 43).

Normen: Es liegen drei bildungsspezifisch (gesamt, ohne und mit Abitur) und sechs altersspezifisch (gesamt, 14-17, 18, 19-22, 23-27 und älter als 28 Jahre) aufgeteilte, also insgesamt 18 Normgruppentabellen vor (Standardwerte $M = 100$, $SD = 10$). Die Größe der Stichproben schwankt je nach Subtest zwischen 2234 und 10024 (Kersting et al., 2008, S. 43).

6.2.2. Multidimensionale Selbstwertkala, Leistungsbezogene Selbstwertschätzung

Die Multidimensionale Selbstwertkala (MSWS) ist eine umfassende Selbstbeschreibungsskala, die die verschiedenen Facetten der Selbstwertschätzung erfasst. Sie basiert auf theoretischen Modellvorstellungen und Ergebnissen empirischer Überprüfungen. Grundlage ist das hierarchische Mehr-Facetten-Modell von Shavelson, Hubner und Stanton (1976), aus dem die Multidimensional Self-Concept Scale (Fleming & Courtney, 1984) entwickelt wurde. Die MSWS ist die adaptierte und modifizierte deutschsprachige Version dieser Skala. Die Selbstwertschätzung hat, unter anderem, zentrale Bedeutung im Umgang mit Belastungen, für die eigene Leistungsfähigkeit, in der Selbstregulation und für Erfolg in den verschiedensten Lebensbereichen. Der Begriff "Selbstwert" lehnt sich dabei an das englische "self-esteem" an und umfasst sowohl im Alltagsgebrauch positiv konnotierte Aspekte wie etwa "Selbstsicherheit" oder "Selbstachtung" als auch negativ konnotierte wie "Selbstgefälligkeit" oder "Eitelkeit". Die Leistungsbezogene Selbstwertschätzung bewertet die Leistungsfähigkeit im Beruf ebenso wie Überzeugungen bezüglich allgemeiner beruflicher und fachlicher Kompetenzen (Schütz & Sellin, 2006).

Das Verfahren erfüllt folgende Testgütekriterien:

Objektivität: Durch die standardisierte, auf dem Selbstbeschreibungsfragebogen abgedruckte Instruktion ist die Durchführungsobjektivität in hohem Masse gegeben, sofern der Testleiter, wie gefordert, keine weiteren Informationen gibt. Die Auswertungsobjektivität ist bei Beachtung der im Manual beschriebenen Anleitungen sichergestellt (Schütz & Sellin, 2006).

Reliabilität: Die interne Konsistenz (Cronbachs Alpha) der Gesamtskala liegt bei 0.93 (Schütz & Sellin, 2006).

Validität: Die Konstruktvalidität ist durch hohe Interkorrelationen von Gesamtwert, Skalen und Subskalen gut belegt. Die konvergente Validität konnte durch hohe Korrelationen mit der Rosenberg-Skala, den Frankfurter Selbstkonzeptskalen und den Frankfurter Körperkonzeptskalen bestätigt werden. Der Nachweis der diskriminanten Validität wurde durch Korrelation mit dem NEO-Fünf-Faktoren Inventar und einem Fragebogen zur Erfassung sozialer Erwünschtheit erbracht (Schütz & Sellin, 2006).

Normen: Die Normierung basiert auf einer nach Alter und Geschlecht geschichteten bevölkerungsrepräsentativen Stichprobe von 453 Personen im Alter von 14 bis 92 Jahren (Schütz & Sellin, 2006).

6.2.3. Erfassung des Grades der Freiwilligkeit

Um die Bereitwilligkeit, an weiteren Befragungen teilzunehmen, zu ermitteln, war folgende Aussage zu bewerten: „Ich bin grundsätzlich bereit, an Befragungen zur Beantwortung von Forschungsfragen teilzunehmen.“ Die siebenstufige Antwortskala dazu reicht von „trifft nicht zu“ bis „trifft genau zu“. Dieses Antwortformat wird auch in der „Multidimensionalen Selbstwertkala“ verwendet. Damit sich die Testpersonen auf ein einheitliches Antwortformat einstellen können, wurde es auch für diese Forschungsfrage verwendet. Zu dem Antwortformat ist anzumerken, dass

sich bei siebenstufigen Ratingskalen extreme Urteile stärker bemerkbar machen und die mittlere Antwortkategorie eher gewählt wird, weil die Testperson die Antwort nicht weiß, diese nicht für wichtig hält oder aus Unsicherheit oder aus Protest gegen die Fragestellung (Kubinger & Jäger, 2003).

6.2.4. AKZEPT! Fragebogen

Der Akzept! Fragebogen zur Messung der Akzeptanz diagnostischer Verfahren ist ein mehrdimensionales Akzeptanz-Messinstrument. Er liegt in vier Varianten vor: Messung der Akzeptanz von Assessment Center, von Interviews, von Persönlichkeitstest und Messung der Akzeptanz von Leistungstests. Für die vorliegende Studie wurde die Version für Leistungstests eingesetzt.

Für die Bearbeitung der Items wird eine sechsstufige Zustimmungsskala genutzt, deren Extremausprägungen mit "trifft nicht zu" und "trifft genau zu" bezeichnet sind. Bei der Auswertung wird ein extrem negativer Wert mit "1", ein extrem positiver Wert mit "6" codiert. Im Folgenden werden die einzelnen Skalen anhand von Beispielitems erläutert:

- Kontrollierbarkeit: "Bei der Bearbeitung der Testaufgaben wusste ich jederzeit, was ich tun muss."
- Messqualität: "Das Interview ermöglicht es, die zwischen verschiedenen Menschen bestehenden Unterschiede in den vom Interview gemessenen Merkmalen exakt zu messen."
- Augenscheinvalidität für Berufseignungsdiagnostik: "Dass man mit den Fragen/Aussagen wie denen des Verfahrens geeignete Personen für einen Job herausfinden kann, ist zu bezweifeln."
- Belastungsfreiheit: "Die Teilnahme ist belastend."
- Gesamtbeurteilung: "Welche Schulnote würden Sie dem soeben bearbeiteten Test geben?"
- Selbsteinschätzung: "Im Vergleich mit anderen Personen meiner Altersgruppe (mit gleicher Schulbildung) denke ich, dass ich im Test ... abgeschnitten habe."

6.3. Durchführung der Untersuchung

Im ersten Teil der Untersuchung werden von allen Teilnehmern, die für diese Studie relevanten Testergebnisse erhoben. Im zweiten Teil der Untersuchung erfolgt die Unterscheidung in Responder und Non-Responder: alle Teilnehmer aus der Testung werden um ihre Teilnahme an einer weiteren Befragung gebeten. Alle, die an dieser teilnehmen, werden zu den Respondern gezählt. Personen, die dies nicht tun, werden als Non-Responder eingestuft.

Um noch vor der Trennung der Teilnehmer in Responder und Non-Responder von beiden Gruppen die Fähigkeit im schlussfolgernden, sprachlichen und räumlichen Denken einheitlich erfassen zu können, konzentrierte sich die Suche nach Teilnehmern auf Gesellschaftsbereiche, in denen Gruppen von Personen gemeinsam getestet werden können. Gleichzeitig soll die Testung für die Teilnehmer auch einen Nutzen bringen, um ihre Bereitschaft für weitere Befragungen nicht zu schwächen. Für Schüler ist die Bearbeitung von Aufgaben, die in ähnlicher Form auch in Bewerbungstests vorkommen können, von Interesse. Eine Teilnahme an den Leistungstests gibt ihnen zudem die Möglichkeit, ihr eigenes Leistungsniveau differenzierter kennen zu lernen. Somit fiel die Entscheidung, Schulklassen, bevorzugt Maturaklassen, um Mithilfe bei der Studie zu bitten.

Im Herbst 2014 erfolgte die Anfrage an die Direktoren der Schulen und den Landesschulrat Niederösterreich. Nach erfolgter Bewilligung und Zustimmung der Schüler selbst zur Teilnahme an der Studie, konnten Termine vereinbart werden. Alle Testungen fanden an einem Vormittag zu Wochenbeginn statt und wurden von derselben Testleiterin (der Studienautorin) durchgeführt, so dass ein einheitliches Vorgehen bei den Testungen in den verschiedenen Klassen garantiert war.

Zu Beginn jeder Testung erfolgte eine kurze Vorstellung. Als Zweck der Befragung wurde eine Untersuchung zur Akzeptanz der durchgeführten Verfahren angegeben, da das Verhalten von Respondern dahingehend interpretiert werden kann, dass ihr Verhalten auf einer höheren Akzeptanz gegenüber Befragungen beruht. Anschließend wurden die Testhefte mit den Antwortbögen und einem leeren Blatt für Notizen entsprechend dem Manual des WIT-2 ausgeteilt. Die Testung erfolgte nach den Anweisungen im Manual: nach dem gemeinsamen Lesen der

Arbeitsanleitung und der Übungsbeispiele wurde, wenn es keine weiteren Fragen mehr gab, die Stoppuhr gestartet und die Teilnehmer bearbeiteten die Aufgaben bis zur „Stopp“ - Markierung des jeweiligen Untertests. Die Vorgabe des Leistungsbezogenen Selbstwertfragebogens erfolgte ohne Zeitangabe. Zusätzlich zu der Frage nach der Bereitschaft an weiteren Befragungen teilzunehmen, waren noch ein paar allgemeine Fragen zur Akzeptanz der Verfahren zu beantworten, um den angegebenen Zweck der Studie für die Teilnehmer nachvollziehbar zu machen. Mit den demographischen Daten wurde auch um eine E-Mail-Adresse gebeten, um die Ergebnisse rückmelden zu können.

Nach dem Absammeln der Testbögen wurden die Teilnehmer darüber informiert, dass mit den Testergebnissen der Akzept!Fragebogen von Kersting zugesendet wird. Es erfolgte die Bitte, auch noch diesen zu bearbeiten, um weitere, wichtige Resonanz zu den Verfahren zu gewinnen.

Die Ergebnisse wurden laut Manual ausgewertet und am Wochenende (fünf Tage nach der Testung) rückgemeldet. Mit dem Ziel, die Rate der Responder durch möglichst geringen Aufwand zu erhöhen, wurde der Fragebogen nicht als Anhang mit gesendet, sondern direkt nach der Bitte um Mithilfe in das E-Mail eingefügt, so dass die Teilnehmer nur auf „antworten“ und, nach Bearbeitung des Bogens, auf „senden“ drücken mussten.

Um den Rücklauf zu erhöhen, erhielten alle Teilnehmer, die bis dahin noch nicht rückgemeldet hatten, nach einer Woche ein zweites Mal den Akzept!Fragebogen zugesendet mit der Bitte, diesen ausgefüllt rückzusenden.

Alle Teilnehmer, die entweder schon beim ersten Mal, oder erst beim zweiten Mal den Fragebogen bearbeitet und rückgesandt haben, werden zu der Gruppe der Responder gezählt. Es erfolgt keine weitere Aufteilung, da die Gründe, warum erst nach der zweiten Aufforderung rückgemeldet wurde, nicht bekannt sind.

6.4. Stichprobe

6.4.1. Stichprobenumfang

Die Berechnung der benötigten Anzahl an Testteilnehmern erfolgt mit Hilfe von G*Power. Es handelt sich dabei um ein Programm zur Durchführung von statistischen Power Analysen von Faul, Erdfelder, Lang und Buchner (2007). Das Programm ist urheberrechtlich geschützt, aber kostenlos für jedermann.

Es sollen die Mittelwerte der Testergebnisse beider Gruppen (Responder vs. Non-Responder) auf signifikante Unterschiede untersucht werden. Da jeder Teilnehmer drei Untertests bearbeitet hat, sind drei Mittelwertsvergleiche nötig. Der Einfluss der Kovariate „leistungsbezogener Selbstwert“ auf die Leistungstestergebnisse soll ausgeschaltet werden. Zu diesem Zweck wird eine einfache multivariate Kovarianzanalyse durchgeführt.

G*Power verlangt zur Berechnung der Stichprobengröße bei der Anwendung einer einfachen multivariaten Kovarianzanalyse die Angabe der angestrebten Effektstärke. Das Programm arbeitet bei Anwendung eines F-Testes mit dem Effektstärkenmaß Cohens f^2 . Dieses beschreibt den Anteil der systematisch durch den Prädiktor aufgeklärten Varianz des Kriteriums gegenüber der Residualvarianz und lässt sich von dem Effektmaß Cohens d ableiten. Geht man davon aus, dass eine Differenz von 12 Intelligenzquotient-Einheiten als in der Praxis bedeutsamer Unterschied zwischen zwei Gruppen zu sehen ist und bezieht diese Differenz auf die Standardabweichung von 15 ergibt sich ein Effekt von Cohens $d = 0,8$. Unter der Annahme, dass in der Stichprobe gleich viele Responder wie Non-Responder sein werden, entspricht Cohens $d = 0,8$ einem Cohens f^2 von 0,16 (siehe Anhang 1).

Das Risiko 1. Art wird mit $\alpha=0,05$ festgelegt. Das Risiko 2. Art wird in vergleichbaren Studien häufig mit $\beta=0,20$ festgelegt. Hier erfolgt eine Festlegung von $\beta=0,15$. Das ergibt eine Wahrscheinlichkeit von 85% für das Erkennen eines tatsächlichen Unterschiedes zwischen Respondern und Non-Respondern, ein für diese Studie zweckmäßiger Wert.

Es ergibt sich somit für diese Untersuchung eine angestrebte Stichprobengröße von 96 Teilnehmern.

6.4.2. Stichprobenzusammensetzung

An der Testung nahmen letztendlich insgesamt 105 Maturantinnen und Maturanten aus fünf niederösterreichischen Klassen teil: 22 Maturantinnen einer Höheren Bundeslehranstalt, eine Abschlussklasse einer HBLA mit 8 weiblichen und 24 männlichen Maturanten, sowie drei Klassen eines Gymnasiums mit insgesamt 25 weiblichen und 26 männlichen Schülern. Das Alter der Schüler variierte zum Zeitpunkt der Testung zwischen 18 und 20 Jahren, wobei 77 % der Teilnehmer 19 Jahre alt waren und somit auch der Altersdurchschnitt bei 19 Jahren lag.

Auch die Verteilung der Geschlechter war relativ ausgeglichen: 52,4% (55 TN) der Teilnehmer waren weiblich, 47,6 % (50 TN) männlich.

7. Ergebnisse

7.1. Auswertung

Die Auswertung erfolgte mit der Statistiksoftware SPSS für Windows (Version 23). Der Einfluss des Faktors Responder/Non-Responder auf die Ergebnisse der drei Leistungstests soll geprüft werden, wobei der Einfluss der Kovariable beseitigt werden soll. Dazu wurde jeweils eine einfache multivariate Kovarianzanalyse durchgeführt. Das Signifikanzniveau wurde auf $\alpha=0,05$ festgelegt.

Abschließend werden die Ergebnisse der Angaben zur Teilnahmebereitschaft an weiteren Befragungen dargestellt.

Bei den durchgeführten Verfahren kann ein Intervallskalenniveau der Daten angenommen werden.

An Voraussetzungen ist die Homogenität der Varianz-Kovarianz-Matrix erforderlich. Diese wurde mittels Box-M-Test geprüft, welcher nicht signifikant ist (siehe Tabelle 1). Die Homogenität der Kovarianz-Matrix ist somit gegeben.

Tabelle 1: Box-M-Test zur multivariaten Analyse der Varianz

Box-M-Test	6,753
F	1,081
df1	6
df2	23498,125
Signifikanz	,371

7.2. Ergebnisse: durch Unterscheidung von Respondern und Non-Respondern ermittelter Effekt

Aus den Mittelwertdifferenzen der einzelnen Leistungstests von Respondern und Non-Respondern dividiert durch die jeweilige Standardabweichung der Variablen, wurde die Effektgröße d nach Cohen ermittelt (siehe Tabelle 2).

Tabelle 2: Cohens d ermittelt aus Mittelwertunterschied dividiert durch Standardabweichung

		Gruppengröße	Mittelwert der Testergebnisse	Standardabweichung	Cohens d
Analogien	Non-Responder	73	109,14	7,158	-0,08
	Responder	32	108,53	8,599	KI [-0.496 - 0.336]
Abwicklungen	Non-Responder	73	105,81	10,173	-0.115
	Responder	32	104,69	8,716	KI [-0.531 - 0.301]
Zahlenreihen	Non-Responder	73	101,86	10,555	-0.01
	Responder	32	101,75	11,906	KI [-0.426 - 0.406]

Da alle Konfidenzintervalle (KI) den Wert Null enthalten, besteht bei keinem der drei Leistungstests ein Effekt bezüglich der Leistungstestergebnisse durch die Unterscheidung von Respondern und Non-Respondern.

7.3. Ergebnisse: einfache multivariate Kovarianzanalyse

Die Durchführung der einfachen multivariaten Kovarianzanalyse brachte folgende Ergebnisse:

Es zeigt sich kein signifikanter Einfluss des Faktors Responder/Non-Responder auf die Leistungstestergebnisse. Die Kovariate „leistungsbezogener Selbstwert“ zeigt keinen Einfluss auf die Leistungstestergebnisse.

Das multivariate Wirkungsausmaß, die erklärte Varianz (partielles Eta-Quadrat) ist bei beiden gering (siehe Tabelle 3).

Tabelle 3: Ergebnisse der multivariaten Kovarianzanalyse

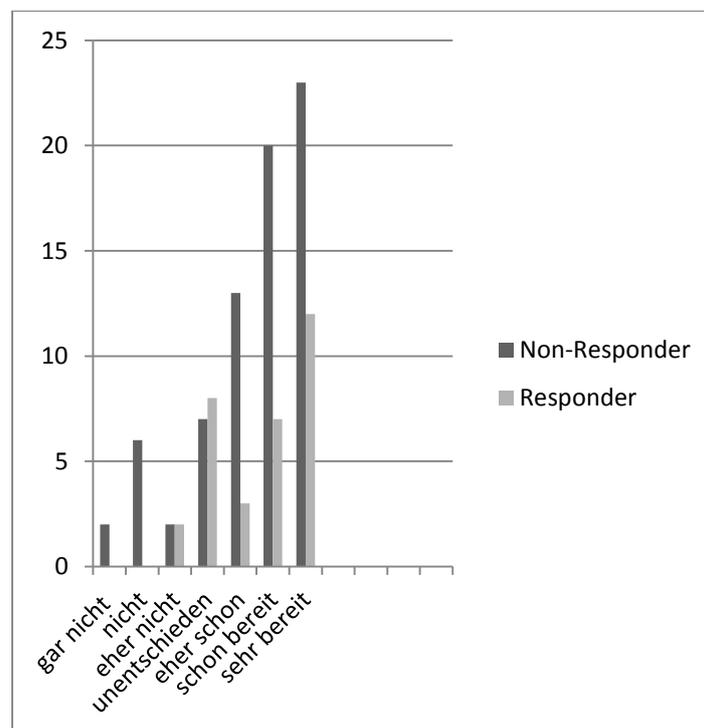
Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz	partielles Eta-Quadrat
Responder/ Non-Responder	Hotelling-Spur	,003	,116	3,000	100,000	,951	,003
leistungsbez. Selbstwert	Hotelling-Spur	,029	,979	3,000	100,000	,406	,029

Von den vier Prüfstatistiken ist hier die Hotelling-Spur als die aufgeklärte Varianz durch die Fehlervarianz (Pillai/Wilks) angeführt.

7.4. Ergebnisse zum angegebenen Grad der Freiwilligkeit

Die Mehrheit aller Studienteilnehmer zeigt sich bereit an weiteren Befragungen teilzunehmen (siehe Abbildung 1).

Abbildung 1: Häufigkeit der Angaben zur Bereitschaft an weiteren Studien teilzunehmen



7.4.1. Ergebnisse zum Grad der Freiwilligkeit der Non-Responder

In dieser Untersuchung gaben von allen Non-Respondern 76,7 % an, „eher schon“ bis „sehr bereit“ zu sein, an weiteren Befragungen teilzunehmen, während nur 13,7 Prozent angaben, „gar nicht“ bis „eher nicht“ teilnehmen wollen. Nur 9,6% zeigten sich unentschieden bezüglich ihrer Bereitschaft zu weiterer Teilnahmen (siehe Tabelle 4).

Tabelle 4: Angaben der **Non-Responder** zu weiterer Teilnahmebereitschaft

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	gar nicht	2	2,7	2,7	2,7
	nicht	6	8,2	8,2	11,0
	eher nicht	2	2,7	2,7	13,7
	unentschieden	7	9,6	9,6	23,3
	eher schon	13	17,8	17,8	41,1
	schon bereit	20	27,4	27,4	68,5
	sehr bereit	23	31,5	31,5	100,0
	Gesamtsumme	73	100,0	100,0	

7.4.2. Ergebnisse zum Grad der Freiwilligkeit der Responder

Von den tatsächlichen Respondern gaben 68,8% an „eher schon“ bis „sehr bereit“ zu sein, weitere Studien zu unterstützen. Die Antwortmöglichkeiten „gar nicht“ und „nicht“ wurden von Respondern dieser Studie nicht gewählt, die Option „eher nicht“ bereit teilzunehmen nur von 6,3 Prozent. 25% zeigten sich unentschieden, an weiteren Befragungen teilzunehmen (siehe Tabelle 5).

Tabelle 5: Angaben der Responder zu weiterer Teilnahmebereitschaft

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	eher nicht	2	6,3	6,3	6,3
	unentschieden	8	25,0	25,0	31,3
	eher schon	3	9,4	9,4	40,6
	schon bereit	7	21,9	21,9	62,5
	sehr bereit	12	37,5	37,5	100,0
	Gesamtsumme	32	100,0	100,0	

8. Diskussion und Ausblick

Das Ziel dieser Arbeit war die Untersuchung möglicher Unterschiede in Bezug auf Intelligenzleistung zwischen Personen, die freiwillig an Befragungen teilnehmen, und sogenannten „Teilnahmeverweigerern“.

Nach Auswertung aller Testergebnisse und erfolgter Zuordnung in die beiden zu vergleichenden Gruppen konnte kein signifikanter Unterschied zwischen Respondern und Non-Respondern in den Leistungstestergebnissen gefunden werden. Auch ein möglicher Einfluss der Kovariate „leistungsbezogenes Selbstwertgefühl“ konnte nicht bestätigt werden. Die Erfassung der Bereitschaft an weiteren Befragungen teilzunehmen, zeigte, dass die Non-Responder dieser Studie mehrheitlich passive Non-Responder waren.

Diese Ergebnisse sprechen dafür, durchaus Freiwillige für die Normierung von psychologischen Verfahren zu verwenden. Unter der Annahme, es gäbe keine Intelligenzunterschiede zwischen Personen, die an Befragungen teilnehmen und solchen, die sich verweigern, können auch Studien mit geringeren Ausschöpfungsquoten verallgemeinert werden.

Um die Aussagekraft dieser Arbeit zu erhöhen, wäre eine dritte Befragung vor Ort zu den Gründen der Verweigerung einer weiteren Teilnahme erforderlich gewesen. Dies war leider vonseiten der Teilnehmer nicht möglich. Es stellt sich darum die Frage, ob es sich um bewusste Entscheidungen für oder gegen eine weitere Teilnahme handelt. Schumann (2012) weist darauf hin, dass man diesbezüglich in der Regel nur begründete Vermutungen anstellen kann. Erst in abschließenden face-to-face Befragungen wäre es möglich, tatsächliche Beweggründe zu erheben und diese Erkenntnisse in eine Folgestudie zu dem Thema einfließen zu lassen. So kann etwa der Zeitpunkt der Teilnahmeaufforderung ungünstig gewesen sein. Es wurde dafür zwar gezielt der Vormittag am Wochenende gewählt unter der Annahme, dass die Gefahr der Ablenkung zu diesem Zeitpunkt am geringsten ist. Trotzdem kann nicht sicher davon ausgegangen werden, dass der Zeitpunkt für die Schüler optimal war. Auch die gängige Praxis, E-Mails am Handy zu empfangen, kann die Teilnahme beeinträchtigen, da die Bearbeitung eines Fragebogens am PC doch einfacher und die Wahrscheinlichkeit, dass Schüler zu einem späteren Zeitpunkt auf die Aufforderung zurückkommen, eher gering ist. Aus diesem Grund

erfolgte nach einer Woche eine zweite Aufforderung. Wiederholte Aufforderungen zu Teilnahme können nach Schützenmeister (2002) allerdings dazu führen, dass Betroffene zur Teilnahme „überredet“ werden. Die Frage nach der Bereitschaft, sich erneut befragen zu lassen, wird so unter Umständen als Ausweg aus dieser für den Befragten unangenehmen Situation genutzt.

Weiters wurde bei dieser Arbeit davon ausgegangen, dass ein deutlicher Unterschied in den Intelligenzwerten von 12 IQ-Werten als praktisch relevant angesehen werden kann. Unter der Annahme, dass schon ein geringerer Unterschied für die Non-Response-Problematik bedeutend ist, wäre für die Durchführung einer solchen Untersuchung eine deutlich größere Stichprobe nötig, um eine ausreichende Teststärke zu erzielen. Die Rekrutierung von Schulklassen, insbesondere Maturaklassen für Studien scheint im Moment nicht einfach, insbesondere die Einführung der Zentralmatura nimmt doch mehr Zeit in Anspruch und macht es schwieriger, Termine für Befragungen oder Testungen zu bekommen. Interessant wäre diesbezüglich der Zugang zu größeren Gruppen von Erwachsenen, die einheitlich getestet werden können, und danach über die Reaktion auf die Bitte zur Teilnahme an einer weiteren Befragung als Responder oder Non-Responder erfasst werden können. Die gewählten Leistungstests sollten dabei für die Teilnehmer interessant sein, um keinen nachteiligen Effekt auf eine weitere Teilnahme zu haben. Nach Schützenmeister (2002) dürfte die Ablehnung der weiteren Kooperation häufig das Ergebnis einer negativen Evaluation der vorhergehenden Befragung sein. Möglicherweise war diese für den Befragten zu lang, zu uninteressant, betraf ihn nur wenig, oder die Fragen wurden als zu persönlich empfunden. Größere Gruppen von Teilnehmern lassen sich eventuell im Bereich des Arbeitsmarktservices, der Österreichischen Bundesbahnen oder ähnlich großstrukturierten Arbeitsbereichen finden.

9. Zusammenfassung

Unter der Annahme, dass sich Responder in ihrer Intelligenzleistung von Non-Respondern unterscheiden, ist es problematisch, Verfahren, die durch die Mitarbeit Freiwilliger geeicht sind, zu verallgemeinern. Eine Generalisierung von Studienergebnissen ist unter Nicht-Berücksichtigung möglicher Unterschiede Freiwilliger und Teilnahmeverweigerer kritisch zu betrachten. Diese Arbeit untersucht, ob sich Leistungstestergebnisse von Respondern und Non-Respondern unterscheiden

Um Daten von Respondern und Non-Respondern vergleichen zu können, wurden im ersten Teil dieser Studie allen Teilnehmern je drei Leistungstests vorgelegt. Zusätzlich wurde der „Leistungsbezogene Selbstwert“ erfasst, um mögliche Einflüsse desselben auf die Ergebnisse kontrollieren zu können. Auch die Bereitschaft, an weiteren Befragungen teilzunehmen, wurde erfragt.

Im zweiten Teil erfolgte die Gruppierung in Responder und Non-Responder: Teilnehmer, die einen weiteren, zugesandten Fragebogen bearbeiteten, wurden in die Gruppe der Responder eingeordnet, zu den Non-Respondern wurden alle gezählt, die diesen nicht beantworteten.

An der Testung nahmen Maturantinnen und Maturanten aus fünf niederösterreichischen Schulklassen teil. Das Alter der Schüler variierte zum Zeitpunkt der Testung zwischen 18 und 20 Jahren, wobei 77 % der Teilnehmer 19 Jahre alt waren und somit auch der Altersdurchschnitt bei 19 Jahren lag.

Auch die Verteilung der Geschlechter war relativ ausgeglichen: 52,4% (55 TN) der Teilnehmer waren weiblich, 47,6 % (50 TN) männlich.

Die Hypothesenprüfung erfolgte mittels einfacher, multivariater Kovarianzanalyse. Es konnten keine signifikanten Unterschiede im Leistungsniveau gefunden werden. Auch ein möglicher Einfluss der Kovariate „leistungsbezogenes Selbstwertgefühl“ konnte nicht bestätigt werden. Die Erfassung der Bereitschaft an weiteren Befragungen teilzunehmen, zeigte, dass die Non-Responder dieser Studie mehrheitlich passive Non-Responder waren.

Bei weiteren Untersuchungen sollte die Möglichkeit bestehen, auch die tatsächlichen Gründe der Nichtbearbeitung des zweiten Fragebogens zu erfassen.

Bei vorhandenen Ressourcen kann mittels größerer Stichproben untersucht werden, ob sich nicht doch ein kleiner Effekt im Leistungsbereich finden lässt.

9.Literatur

- Aronson, E., Wilson, T.D. & Akert R. (2008). Sozialpsychologie.6., aktualisierte Auflage. München: Pearson Studium.
- Benölken, R. (2014). Begabung, Geschlecht und Motivation: Erkenntnisse zur Bedeutung von Selbstkonzept, Attribution und Interessen als Bedingungsfaktoren für die Identifikation mathematischer Begabungen. *Journal für Mathematik-Didaktik*, 35:129-158. doi: 10.1007/s13138-013-0059-9
- Burkell,J. (2003). The dilemma of survey nonresponse. *Library & Information Science Research*, 25, 239-263.
- Carlo, G., Okun, M.A., Knight, G.P. & de Guzmán, M.R.T. (2005). The interplay of traits and motives on volunteering: agreeableness, extraversion and prosocial value motivation. *Personality and Individual Differences*, 38, 1293-1305.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. (pdf)
- Fleming, J.S. & Courtney, B.E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology*, 46, 404-421.
- Gottschling, J. (2013). Eine Multi-Kohorten-Sequenz-Analyse des Zusammenhangs von Motivation und Schulerfolg unter Berücksichtigung allgemeiner kognitiver Fähigkeiten. Ergebnisse aus der Zwillingsstudie KoSMoS. Dissertation; *Literature* (61; 99). Universität Saarbrücken.
- Haunberger, S. (2011). *Teilnahmeverweigerung in Panelstudien*. 1. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Kersting, M., Althoff, K. & Jaeger, A.O. (2008). WIT-2. Wilde-Intelligenz-Test 2. Goettingen: Hogrefe.
- Kubinger, K.D. & Jäger, R.S. (2003). Schlüsselbegriffe der Psychologischen Diagnostik. Berlin: Beltz.
- Kurz, H. (1987). Die Genauigkeit von Umfrageergebnissen, eine empirische Analyse des Nonresponse-Bias. Service Fachverlag an der Wirtschaftsuniversität Wien.
- Lenhard, W. & Lenhard, A. (2015). Berechnung von Effektstärken. Verfügbar unter: <http://www.psychometrica.de/effektstaerke.html>. Bibergau: Psychometrica.
- Lyness, K. S. & Brumit Kropf, M. (2007). Cultural values and potential nonresponse bias. A multilevel examination of cross-national differences in mail survey response rates. *Organizational Research Methods*, 10, 210-224.
- Lonnqvist, J. E., Paunonen, S., Verkasalo, M., Leikas, S., Tuulio-Henriksson, A. & Lonnqvist, J. (2007). Personality characteristics of research volunteers. *European Journal of Personality*, 21 (8), 1017-1030.
- Neller K. (2005). Kooperation und Verweigerung: Eine Non-Response-Studie. *ZUMA Nachrichten*, 57, 9-36.
- Marcus, B. & Schütz, A. (2005). Who Are the People Reluctant to Participate in Research? Personality Correlates of Four Different Types of Nonresponse as Inferred from Self- and Observer Ratings. *Germany Journal of Personality* 73:4. doi: 10.1111/j.1467-6494.2005.00335.x
- Porter, S.R. & Whitcomb, M. E. (2005). Nonresponse in student surveys: The role of demographics, engagement and personality. *Research in Higher Education*, 46, 127-152.

- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Berlin, Heidelberg : Springer Berlin Heidelberg.
- Rogelberg, S.G., Conway, J.M., Sederburg, M.E., Spitzmüller, C., Atiz, S. & Knight, W.E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, 88 (6), 1104-1114.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen : Leske + Budrich.
- Schräpler, J.-P. (2000). Was kann man am Beispiel des SOEP bezüglich Nonresponse lernen? *ZUMA-Nachrichten* 46, 117-149.
- Schumann, S. (2012). *Repräsentative Umfrage: Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren*. Berlin: Walter de Gruyter.
- Schütz, A. & Sellin, I. (2006). *Multidimensionale Selbstwertskala*. Göttingen: Hogrefe.
- Schützenmeister, F. (2002). Die Bereitschaft, sich wieder befragen zu lassen, in postalischen Erhebungen *Zeitschrift für Soziologie*, Jg. 31, Heft 2, , S. 138–154.
- Shavelson, R.J., Hubner, J.J. & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46 (3), 407-441.
- Stockè, V. (2006). Attitudes toward Surveys, Attitude Accessibility and the Effect on Respondents' Susceptibility to Nonresponse. In *Quality & Quantity* 40:259–288 Springer: University of Mannheim.

Tiapal, A. (2011). Der Einfluss von Persönlichkeit und Motivdispositionen auf Nonresponse. Diplomarbeit, Universität Wien. Fakultät für Psychologie.

Unterhofer, E. (2009). Eine Responder-Analyse zur Erfassung des Einflusses der Freiwilligkeit auf die Testeichung und – validierung. Diplomarbeit, Universität Wien. Fakultät für Psychologie.

Wentura, D. (2004). Ein kleiner Leitfaden zur Teststärken-Analysen. Universität Saarland. Fakultät Psychologie. Verfügbar unter:
<http://www.researchgate.net/publication/242433858>

Wiesner, K. (2011). Die Bedeutung des Impostor-Phänomens bei jungen WissenschaftlerInnen. Diplomarbeit, Universität Wien. Fakultät für Psychologie.

Anhang 1

Ableitung des Effektstärkenmaßes Cohens f^2 aus Cohens d für die Anwendung von G*Power zur Bestimmung des Stichprobenumfang:

$$a = (n_1+n_2)^2 / n_1 n_2 \quad a = (50+50)^2 / 50*50 = 4$$

$$r = d / \sqrt{d^2 + a}$$

$$r = \mathbf{0,8} / \sqrt{0,64 + 4} = 0,37$$

$$f^2 = r^2 / (1 - r^2)$$

$$f^2 = 0,138 / (1 - 0,138) = \mathbf{0,16}$$

Anhang 2

Elternbrief

Liebe Eltern,

Ihre Kinder bekommen im Rahmen des Psychologieunterrichts die Möglichkeit zur Teilnahme an einem Testverfahren und somit folgendes Angebot:

- ❖ Kennenlernen der Situation bei Aufnahmetests und Personalausleseverfahren, mit welchen sie möglicherweise auf ihrem weiteren Bildungsweg konfrontiert werden
- ❖ Kennenlernen typischer, bei Leistungstests immer wieder vorkommender Fragen
- ❖ bessere Einschätzung der eigenen Stärken

Der gewählte Test *WIT-2* erlaubt eine detaillierte Vorhersage des Erfolgs in Schule, Ausbildung, Studium und Beruf.

Ergebnisse werden per E-Mail an die jeweiligen Teilnehmer direkt zugesandt: weder Lehrer noch Mitschüler erfahren Testergebnisse anderer.

Die Teilnahmemöglichkeit ergibt sich aus der Durchführung einer Diplomarbeit zur Akzeptanz von Leistungstests an der Universität Wien. Als Diplomandin der Studienrichtung Psychologie und als Testleiterin werde ich selbstverständlich alle Daten vertraulich behandeln.

Mit der Bitte um Ihr Einverständnis zur Teilnahme Ihrer Tochter / Ihres Sohnes

Goll Ulrike

Meine Tochter /mein Sohn _____ darf im Rahmen des Psychologieunterrichtes an dem Testverfahren teilnehmen.

Unterschrift des Erziehungsberechtigten

Anhang 3

Aufforderung zu weiterer Teilnahme zwecks Unterscheidung von Respondern und Non-Respondern

Liebe Testteilnehmer, liebe Testteilnehmerinnen,

Wie vereinbart schicke ich Euch Eure Ergebnisse mit ein paar kurzen Fragen, wie zufrieden Ihr mit dem Test wart, zu.

Bitte gleich auf ANTWORTEN klicken – Kreuzerl in die zutreffenden Kasterl machen (dauert keine 2 Minuten) und SENDEN:

AKZEPT! Bogen:.....

VIELEN DANK!

Ihr helft mir damit sehr beim Schreiben meiner Diplomarbeit, bei der ich mich ja damit befasse, wie annehmbar Testungen sind.

Nochmals herzlichen Dank und ALLES GUTE für Eure Matura!

Mit lieben Grüßen

Ulrike Goll

Anhang 4

Rückmeldung der Ergebnisse

Am Testtag, dem 23.12.2014 haben Sie folgende Werte erreicht (entspricht der Tagesverfassung):

Der Standardwert wird aus einer Vergleichsstichprobe gewonnen, um individuelle Testergebnisse einordnen zu können. Der Durchschnitt der Bevölkerung erreicht zwischen 90 und 110 Standardwertpunkte.

Der Prozentrang gibt an, wie viel Prozent der Vergleichsgruppe eine bessere, gleich gute oder schwächere Leistung erreichen. Zum Beispiel besagt ein Prozentrang 40, dass von hundert vergleichbaren Individuen 60 eine gleich gute bzw. bessere und 40 eine schwächere Leistung erzielen.

Durchgeführte Subtests	Aufgabe	erfasste Fähigkeit	Ergebnis
Analogien	Auf der linken Seite eines Gleichheitszeichens sind zwei Wörter vorgegeben, die in einer bestimmten Beziehung zueinander stehen, auf der rechten Seite ein Wort. Von fünf Wahlwörtern ist dasjenige auszuwählen, das auf der rechten Seite des Gleichheitszeichens eine analoge Beziehung herstellt.	Die Fähigkeit, mit sprachlichen Konzepten umzugehen, wobei Wortschatz, Sprachverständnis und sprachlogisches Denken eine Rolle spielen.	Standardwert: <u>xxx</u> Prozentrang: <u>xxx</u>
Zahlenreihen	Vorgegeben ist eine Folge von Zahlen, die nach einer Regel aufgebaut ist; auf dem Antwortbogen sind die Ziffern der Zahl durchzustreichen, die als nächstes Glied der Reihe folgen müsste	Die Fähigkeit, bestimmte logische Regeln und Gesetzmäßigkeiten zu erkennen und zweckentsprechend anwenden zu können	Standardwert: <u>xxx</u> Prozentrang: <u>xxx</u>
Abwicklungen	Zu einer aus mehreren Flächen mit verschiedenen Zeichen bestehenden Faltvorlage ist aus fünf vorgegebenen Körpern derjenige herauszufinden, der sich aus der Faltvorlage herstellen lässt	Die Fähigkeit zur Vorstellung räumlicher Relationen und bestimmte logische Regeln und Gesetzmäßigkeiten zu erkennen und zweckentsprechend anwenden zu können.	Standardwert: <u>xxx</u> Prozentrang: <u>xxx</u>

Lebenslauf

Persönliche Angaben:

Name: Goll Ulrike
 E-Mail: buschenschank-goll@aon.at
 Geburtsdaten: 22.02.1966, Tulln
 Staatsangehörigkeit: Österreich
 Familienstand: verheiratet, 3 Kinder

Schulische Ausbildung:

1972 – 1976 Volksschule Hausleiten
 1976 – 1980 Realgymnasium Stockerau
 1980 – 1985 HBLA Sitzenberg-Reidling

Studium:

1985 Inskription an Universität für Bodenkultur
 1991 Diplomarbeit (Pflanzenschutz im Knoblauchanbau)
 1996 Diplomprüfung und Sponsion

 2010 Anmeldung zum Psychologiestudium
 2012 1. Diplomprüfung
 2015 Diplomarbeit im Bereich Psychologische Diagnostik

Berufliche Tätigkeit:

1991 Führung eines landwirtschaftlichen Betriebes gemeinsam mit Ehemann
 1997 Planung und Realisierung eines Buschenschankbetriebes
 1998 Erster Ausschank
 2005 Zertifikationslehrgang für Seminarbäuerinnen
 2006 – 2009 Kursleiterin in Erwachsenenbildung

Praktika:

Zwischen 1983 und 1990 diverse Ferialpraktika:
 landwirtsch. Betrieb mit Milchviehhaltung
 landwirtsch. Betrieb mit Mohnanbau und
 Tockenblumenkulturen
 Betreuerin in Kindererholungslagern
 Service in Hotel im Ötztal
 Zuck erforschungslabor Tulln
 Saatgutlabor Korneuburg

Hobbies: Organistin, Klavier, Bücher