

Christof Nachtigall & Andreas Wolf

Fragebogen zur
Wahrscheinlichkeitstheorie (FWT)



Impressum

methevalreport
erscheint seit 1999
in unregelmäßigen Abständen
als „graue“ Schriftenreihe des Lehrstuhls für
Psychologische Methodenlehre und Evaluationsforschung
am Institut für Psychologie der Friedrich-Schiller-Universität Jena

Herausgeber:
Prof. Dr. Rolf Steyer
Skr.: + 49 (3641) 945 230
Durchwahl: + 49 (3641) 945 231
Fax: + 49 (3641) 945 232

rolf.steyer@uni-jena.de

Redaktion:
Dipl.Psych. Friedrich Funke
sff@uni-jena.de

Typographie:
cand.psych. Silke Zachariae
zachariae@web.de

Standort:
Thüringer Universitäts- und Landesbibliothek
Lesesaal Zweigstelle Psychologie

Internet
<http://www.uni-jena.de/sw/metheval/report/>

Bestellungen:
Methodenlehre und Evaluationsforschung
Institut für Psychologie
Steiger 3 Haus 1
D-07743 Jena
Deutschland

Copyright:
Bei unveröffentlichten Arbeiten verbleibt das Urheberrecht bei der Autorin oder beim Autor.
Das Copyright für Texte, die in anderen Publikationsorganen erschienen sind, liegt bei diesen Organen.

Fragebogen zur Wahrscheinlichkeitstheorie (FWT)

von

Christof Nachtigall¹ und Andreas Wolf

Handanweisung

¹ Korrespondenzadresse: Dr. Christof Nachtigall
Institut für Psychologie
Friedrich-Schiller-Universität Jena
Steiger 3 Haus 1, 07743 Jena
christof.nachtigall@uni-jena.de

Vorwort

Diese Handanweisung zum Fragebogen zur Wahrscheinlichkeitstheorie (FWT) soll dem Nutzer die notwendigen Informationen zur Anwendung dieses Instrumentes, zu seinem theoretischen Hintergrund, seinen testtheoretischen Gütekriterien (Reliabilität, Validität) und zu seiner Konstruktion geben.

Der FWT wurde im Rahmen der Jenaer Panel-Studie zu Lernzuwächsen im Fach Wahrscheinlichkeitstheorie entwickelt (Nachtigall, Kämpfe, Kärtner, Kraus, Preißler & Wolf, 2001). Ziel des Projekts war es, den Wissensstand und die Veränderung des Wissensstandes in möglichst einfacher und ökonomischer Weise zu erfassen. Gleichzeitig sollte der Test die Verwendung latenter Variablen ermöglichen. Für diese Zwecke, insbesondere für die Modellierung der "wahren Veränderung" mit sogenannten True-Change-Modellen (siehe Steyer, Eid & Schwenkmezger, 1997; Nachtigall, Kraus & Steyer, 2000) und die Prädiktion dieser Veränderung durch Drittvariablen stellt der FWT ein besonders geeignetes Instrument dar.

Dank gebührt den Kollegen und Studierenden verschiedener deutscher Universitäten, die im Rahmen der mehrjährigen Jenaer Panel-Studie an den Datenerhebungen teilnahmen und deren Kooperation die Grundlage für die Entwicklung dieses Tests und der darauf basierenden Analysen zu Lernfortschritten im Fach WT bilden.

Christof Nachtigall

Andreas Wolf

Inhaltsverzeichnis

Vorwort	1
Teil I – Anwendung	3
1. Wozu dient der FWT?	3
2. Zur Wahrscheinlichkeitstheorie	3
3. Die Items	3
4. Darbietungsformen	4
5. Durchführung	4
6. Auswertung	5
7. Interpretation	6
Teil II – Statistische Analysen und Gütekriterien	7
1. Beschreibung der Datenerhebung	7
2. Verteilungskennwerte und Normen	8
3. Gütekriterien	8
4. Itemanalyse	10
5. Veränderungsmessung mit dem FWT	12
Literatur	13
Anhang	14
A. Prozentränge und Verteilung der Testscores	14
B. Korrelationen aller Items zu jedem der 3 Messzeitpunkte	16
C. Die Items des FWT	18
D. Prozentränge und Verteilungen der Differenzenscores	19

Teil I - Anwendung

In diesem ersten Teil der Handanweisung soll das Wichtigste zum Verständnis und zur Anwendung des Fragebogen zur Wahrscheinlichkeitstheorie (FWT) dargelegt werden. Dazu gehören eine kurze fachliche Einführung, die Vorstellung der Items sowie die Anleitung zur Durchführung und Auswertung des FWT.

1. Wozu dient der FWT?

Der Fragebogen zur Wahrscheinlichkeitstheorie ist ein Instrument zur Erfassung des Wissens und der Wissensveränderung im Bereich elementarer Stochastik. Mittels 14 Fragen aus 7 verschiedenen Teilgebieten dieses Wissensgebietes wird der zum Messzeitpunkt vorliegende Wissensstand der Testpersonen erhoben.

Anwendungsbereiche des FWT sind neben der Veränderungsmessung in der Forschung die Leistungserfassung von Lernenden/Studierenden und die Evaluation der Wirksamkeit von Trainings in diesem mathematischen Teilgebiet.

2. Wahrscheinlichkeitstheorie

Der Umgang mit dem Konzept 'Wahrscheinlichkeit' ist eine Schlüsselqualifikation für Schüler und Studierende im Rahmen einer wissenschaftlichen Ausbildung. Nicht nur für die Naturwissenschaften, auch für die Human-, Sozial- und Wirtschaftswissenschaften stellt Wahrscheinlichkeitstheorie (WT) ein unverzichtbares Werkzeug dar, welches vor allem in der statistischen Anwendung zum Tragen kommt. Vorlesungen in Methodenlehre und Statistik sind z.B. im Rahmen des Psychologiestudiums an deutschen Universitäten Pflichtveranstaltungen und stellen für viele Studierende eine ernste Hürde auf dem Weg zum Diplom dar (siehe z. B. Schulmeister, 1983).

Die Entwicklung des FWT gründet in diesem Kontext. Ziel war es, ein einfach anzuwendendes Messinstrument für Kenntnisse und Lernzuwächse in der WT zu konstruieren. Der vorliegende Test wurde an einer großen Stichprobe von Psychologiestudierenden entwickelt und erprobt.

3. Die Items

Der FWT enthält 14 Items, welche die wichtigsten Teilgebiete elementarer Stochastik abdecken. Diese Gebiete sind:

- Laplace-Wahrscheinlichkeit (LW)
- Kombinatorik (KO)

- Bedingte Wahrscheinlichkeit (BW)
- Gegenwahrscheinlichkeit (GW)
- Wahrscheinlichkeit stochastisch unabhängiger Ereignisse (UE)
- Spezielle Wahrscheinlichkeitsverteilungen (Binomialverteilung) (BI)
- Erwartungswerte von Zufallsvariablen (EW)

Die Items wurden zumeist als Würfel- oder Münzwurfaufgaben formuliert, bei denen die gesuchte Wahrscheinlichkeit anzugeben ist. Beispiel: Ein Item zur Laplace-Wahrscheinlichkeit lautet:

Item LW1: Wie wahrscheinlich ist es, beim einfachen Würfelwurf eine Augenzahl zu werfen, die mindestens 3 ist? _____

Für jedes der sieben Teilgebiete enthält der FWT zwei parallel formulierte Items. Das parallele Item zur Laplace-Wahrscheinlichkeit lautet:

Item LW2: Wie wahrscheinlich ist es, beim einfachen Würfelwurf eine Augenzahl zu werfen, die höchstens 2 ist? _____

Die Liste aller 14 Items findet sich im Anhang C.

4. Darbietungsformen

Der FWT liegt in zwei Formen (A und B) vor. Diese bestehen aus je 14 Items. Dabei unterscheiden sich die Items von Form A und Form B nicht durch die Art der Aufgabenstellung, wohl aber in deren Reihenfolge und in den konkreten Ereignissen – und damit in ihren Lösungen. Die beiden Formen ermöglichen die Durchführung des FWT als Gruppentest, etwa im Rahmen einer universitären Lehrveranstaltung. Darüber hinaus besteht jede der beiden Formen aus zwei parallelen Testhälften, welche aus je 7 Items bestehen. Mittels dieser Testhälften können die Scores des FWT in einfacher Weise als latente Variablen modelliert werden (siehe z.B. Nachtigall et al., 2000).

Außerdem liegt der FWT neben der Papier und Bleistift Version noch in computerisierter Form vor (Wolf & Nachtigall 2001).

5. Durchführung

Der FWT kann als Einzel- oder Gruppentest durchgeführt werden. Im letzteren Fall erhalten Sitznachbarn jeweils unterschiedliche Formen des Tests. Neben einem ausreichenden Sprachverständnis werden keine weiteren Fähigkeiten oder Kenntnisse vorausgesetzt. Es sollten keine Rechenhilfen oder anderes Material für den Test Verwendung finden. Für die Durchführung des Tests als Papierversion sind nur der Fragebogen, bestehend aus einem DIN A4-Bogen, und Schreibzeug notwendig. Die computerisierte Version des FWT ist webbasiert und benötigt einen Internet/Browser ab Version 4 (Internet Explorer 5.0 oder höher, Netscape Navigator 4.6 oder höher). Die computerisierte Version ist optimiert für den Windows Internet Explorer 5.5.

Dem Fragebogen ist eine Instruktion vorangestellt. Der Versuchsleiter liest diese Instruktion laut vor. Wenn keine Fragen der Probanden mehr bestehen, fordert der Versuchsleiter die Probanden zur schriftlichen Beantwortung der 14 Items auf. Die Antworten werden in freiem Antwortformat auf den dafür vorgesehenen Platz im Fragebogen eingetragen. In der Regel ist die Antwort eine Wahrscheinlichkeit in der Form eines mathematischen Bruchs oder einer Dezimalzahl.

Die maximale Beantwortungszeit beträgt *15 Minuten*. Zusammen mit der Instruktion dauert die gesamte Durchführung etwa 18 Minuten. Obwohl der Fragebogen nicht als Schnelligkeitstest konzipiert ist, wird für die Vergleichbarkeit der Ergebnisse die Einhaltung der Durchführungszeit empfohlen. Der Versuchsleiter erteilt nach 5 Minuten den Hinweis, schwierige Aufgaben zu überspringen und zunächst solche Aufgaben zu lösen, deren Lösung bekannt ist.

6. Auswertung

Die Auswertung erfolgt in der Papierversion durch instruierte Hilfspersonen mit Hilfe einer Auswertungsschablone. Diese gehört zum Materialanlage des Tests (oder kann über die Autoren bestellt werden). Für jede richtige Antwort wird ein Punkt vergeben. Dabei ist zu beachten, dass korrekte Lösungen in verschiedenen Formen angegeben werden dürfen, etwa als gewöhnlicher Bruch oder als Dezimalzahl.

Beispiele:

Als korrekte Lösung des Items LW1 sind die Ausdrücke $4/6$, $2/3$, $0.66\dots$ (mit mindestens 2 Nachkommastellen), 0.67 , 0.667 und $0.\overline{6}$ zu bewerten.

Als korrekte Lösung des Items EW1 (Erwartungswert beim dreimaligen Würfelwurf) ist der exakte Wert 10.5 bzw. $10\frac{1}{2}$, nicht aber ganzzahlige Rundungen oder die Angabe eines Intervalls (10 bis 11) zu bewerten.

Geben die Testpersonen die Lösungen in konsistenter Weise durch einen Wettquotienten $P(A):P(-A)$ an, so sind diese ebenfalls als richtig zu bewerten (wobei $P(A)$ die Wahrscheinlichkeit für das Eintreten des Ereignisses und $P(-A)$ deren Gegenwahrscheinlichkeit darstellt). Bei Item LW1 wäre dies der Ausdruck $4:2$ bzw. $2:1$.

Eine Ausnahme bilden die Items zur Gegenwahrscheinlichkeit (GW1 und GW2), zu deren richtiger Beantwortung die Angabe der Berechnungsformel $1-(5/6)^k$ ausreicht.

Die Punktzahlen der Testhälften bzw. die Gesamtpunktzahl ergibt sich aus der ungewichteten Summe der richtig gelösten Items. In der Auswertungsschablone sind die zu den einzelnen Testhälften gehörenden Items durch eine einheitliche Farbe markiert.

In der computerisierten Fassung erfolgt die Auswertung automatisch. Der Rohpunktwert wird zusammen mit der Testpersonenennung in einer Textdatei gespeichert (CSV – Comma Separated Values oder XML).

7. Interpretation

Der FWT erfasst den momentanen Wissenstand einer Testperson im Bereich elementarer Wahrscheinlichkeitstheorie. Je nach dem Stand im Wissensgebiet sind unterschiedliche Gesamtleistungen der Testpersonen zu erwarten. Eine Normierung der Daten erfolgte an der Population der Erstsemester im Hauptfach Psychologie, und zwar bevor und nachdem der relevante Stoff in der Vorlesung behandelt wurde. Die entsprechenden Prozentränge finden sich in Tabelle 5 im Anhang A.

Die Interpretation der Ergebnisse ist sehr einfach: hohe Punktwerte sind gleichbedeutend mit einem hohen Wissen im Bereich Wahrscheinlichkeitstheorie, niedrige Punktzahlen mit geringem Wissen (siehe Abschnitt II.3 zur Validität des FWT). Was jeweils "hoch" und "niedrig" heißt, kann man an den Verteilungskennwerten wie Mittelwert und Standardabweichung ablesen. Bei Bedarf kann man den Skalenrohwert X in einen Standardwert mit Mittelwert M und Standardabweichung K umrechnen. Das geschieht nach der Formel: $SW = M + K(X - \bar{X})/s$. Dabei ist M der gewünschte Mittelwert der standardisierten Skala, K deren Standardabweichung, X der Skalenrohwert, \bar{X} der Mittelwert und s die Standardabweichung (vgl. Tabelle 1). Darüber hinaus können die Prozentränge der Versuchspersonen direkt in Tabelle 5 im Anhang A abgelesen werden.

Es ist darauf hinzuweisen, dass es in der Population von Erstsemestlern Psychologie zu keinen sogenannten Deckeneffekten kam. Weniger als ein Prozent der Versuchspersonen erreichten bei der 1. Welle den maximalen Punktwert von 14 Punkten im FWT, bei den Wellen 2 und 3 waren es 4 bzw. 2 Prozent der Teilnehmer.

Bei der Untersuchung von *Lernzuwächsen* sind nicht die absoluten Testscores sondern die Veränderung der Scores einer Person im Vergleich zu einem anderen Messzeitpunkt von zentralem Interesse. In den Tabellen 4 und 10 und den Abbildungen 4 – 6 sind die Verteilungen und die Prozentränge der Differenzenscores wiedergegeben. Mit deren Hilfe können individuelle Lernzuwächse zwischen zwei Wellen mit denen der Population verglichen werden (vgl. Abschnitt II.5 über Veränderungsmessung mit dem FWT).

Teil II – Statistische Analysen und Gütekriterien

In diesem zweiten Teil des FWT-Handbuchs werden die Ergebnisse zur statistischen Analyse der testtheoretischen Gütekriterien berichtet, insbesondere zur Schwierigkeit und Trennschärfe der Items sowie zur Reliabilität und Validität des Tests.

1. Beschreibung der Datenerhebung

Der FWT wurden insgesamt 1075 Psychologiestudierenden des 1. Fachsemesters mehrfach vorgelegt. Die Erhebung erfolgte im Zeitraum von 1997-2000 an fünf deutschen Universitäten². Sie wurde als Gruppentest in den Lehrveranstaltungen, in denen Wahrscheinlichkeitstheorie gelehrt wurde, durchgeführt. In der Regel handelte es sich dabei um eine Vorlesung in Statistik bzw. Methodenlehre. Das Panel umfasste jeweils drei Messzeitpunkte: Die erste Erhebung (Welle) fand direkt vor Behandlung des Themas Wahrscheinlichkeitstheorie in der Vorlesung statt, die zweite Erhebung direkt im Anschluss an die Behandlung des Stoffes (im Durchschnitt 4 Wochen später). Die dritte Erhebung erfolgte zu Beginn des 2. Semesters (im Durchschnitt 4 Monate später).

In die folgenden Analysen gehen aus Gründen der Vergleichbarkeit nur die Daten der N=800 Diplomstudierenden ein. Die Daten aus allen Erhebungsorten und Jahrgängen wurden dazu zu einem Datensatz aggregiert. Normwerte und Prozentränge beziehen sich daher auf die Population von Studierenden im Diplomstudiengang Psychologie in Deutschland. An der ersten Welle nahmen N=642 Versuchspersonen teil, an der zweiten Welle N=498 Versuchspersonen, an der dritten Welle N=450 Versuchspersonen. Der Ausfall von Versuchspersonen erklärt sich durch Fluktuation und Schwund der Studierenden in den Lehrveranstaltungen im Laufe eines Semesters bzw. die Nicht-Fortsetzung im nächsten Semester.

Von den 800 erfassten Fragebögen stammen 79% von Frauen und 21% von Männern. Das Alter der Versuchspersonen variierte zwischen 17 und 47 Jahren, der Mittelwert betrug 22.5, der Median 21 Jahre. Alle Teilnehmer besitzen die allgemeine Hochschulreife und sind eingeschriebene Vollstudenten im Fach Psychologie.

² An einer der beteiligten Universitäten wurde das Themengebiet 'Wahrscheinlichkeitstheorie' erst im 2. Semester behandelt (n=126), woraufhin die Messungen jeweils 1 Semester später als bei den übrigen Universitäten erfolgten. Die mittleren Testscores dieser Datenquelle sind jedoch nicht signifikant verschieden von denen anderer Universitäten.

2. Verteilungskennwerte und Normen

In Tabelle 1 dargestellt sind die Mittelwerte und Streuungen der Testscores in den einzelnen Wellen. Im engeren Sinne sind nur die Daten der ersten Welle als Normwerte im Sinne eines Leistungstests zu verstehen. Die Ergebnisse der zweiten und dritten Welle werden z.B. durch Dauer und Intensität der entsprechenden Lehrveranstaltung beeinflusst. Allerdings ist dieser Einfluss mit einem erklärten Varianzanteil von 1,5 % bei der zweiten Welle bzw. 0,6% bei der dritten Welle auf die FWT-Scores gering.

Welle	Stichprobe	Mittelwert	Standard- abweichung	Median
1	642	4,50	3,11	4,00
2	498	7,66	3,26	8,00
3	450	6,93	3,20	7,00

Tabelle 1: Mittelwerte und Standardabweichungen zu den n=3 Messzeitpunkten. Die erste Messung ist vor der Behandlung der Wahrscheinlichkeitstheorie im Rahmen einer Lehrveranstaltung; Messzeitpunkt 2 ist unmittelbar nach der Behandlung des Stoffes (ca. 4 Wochen später); Messzeitpunkt 3 ist 4 Monate nach Abschluss der Lehrveranstaltung.

Differenziertere Angaben über die Verteilung der Testscores in den drei Wellen liefern die Häufigkeitsverteilungen sowie die Tabellen der Prozentränge (siehe Anhang A).

3. Gütekriterien

Der FWT wurde zur objektiven, reliablen und validen Messung der aktuellen Kenntnisse von Testpersonen in elementarer Wahrscheinlichkeitstheorie konzipiert. Die einzelnen Gütekriterien werden im Folgenden diskutiert.

Objektivität

Bei korrekter Auswertung mit Hilfe der Auswertschablone und den in Abschnitt I.6 beschriebenen Auswertungsrichtlinien stellt der FWT ein objektives Messinstrument dar.

Reliabilität

Der FWT besteht aus zwei parallelen Testhälften zu je 7 Items. Die Reliabilität ist in Tabelle 2 abzulesen.

Reliabilität	1. Messung n=642	2. Messung n=498	3. Messung n=450
Testhälften	0,82	0,83	0,85
Gesamttest ³	0,90	0,91	0,92

Tabelle 2: Reliabilität der Messungen in den Wellen 1 - 3.

Die Reliabilität der Testhälften ist mit 0,82 - 0,85 bereits zufriedenstellend. Die Reliabilität des Gesamttestes von 0,90 bis 0,92 kann als gut bis sehr gut bewertet werden.

Bei der Schätzung der Reliabilität wird nicht Cronbach's α verwendet, da die einzelnen Items des FWT nicht im messtheoretischen Sinne als τ -äquivalent gelten können. Dies beruht auf der Tatsache, dass bei der Itemauswahl die Breite der verschiedenen wahrscheinlichkeitstheoretischen Themengebiete berücksichtigt wurde (siehe der folgende Abschnitt über Validität). Lernzuwächse sind aber auch in einzelnen Teilgebieten möglich. In dieser Situation stellt Cronbach's α nur eine untere Schranke der Reliabilität dar (siehe Steyer und Eid, 1996, S. 131).

Auch die Retest-Korrelation wird nicht zur Reliabilitätsschätzung verwendet, da der Test explizit zur Veränderungsmessung konzipiert ist und sich die Merkmalsausprägungen zwischen den Messzeitpunkten verändert.

Validität

Die Items des FWT stammen aus 7 grundlegenden Teilgebieten der Wahrscheinlichkeitstheorie. Diese sind Laplace-Wahrscheinlichkeit, Kombinatorik, bedingte Wahrscheinlichkeit, Gegenwahrscheinlichkeit, Wahrscheinlichkeit stochastisch unabhängiger Ereignisse, spezielle Wahrscheinlichkeitsverteilungen (Binomialverteilung) und Erwartungswerte von Zufallsvariablen. Alle Items beziehen sich auf diskrete Zufallsexperimente; in diesem Sinn ist das zu messende Konstrukt die stochastischen Fähigkeiten der Probanden. Die durch die Items repräsentierten Teilgebiete gehören daher zur elementaren Wahrscheinlichkeitstheorie. Die Themenbreite entspricht den Eingangskapiteln universitärer und schulischer Einführungsliteratur (z.B. Plachky, Baringhaus & Schmitz, 1983; Griesel & Postel, 1990; Feuerpfeil und Heigl 1991; Schmid und Schweizer 1988). Für die Itemformulierung wurden in der Regel die prototypischen Kontexte des Münzwurf- und Würfelexperimentes verwendet, wie

³ gemäß Testverdopplungsformel von Spearman-Brown.

sie in wahrscheinlichkeitstheoretischen Lehrbüchern üblich sind. Daher kann im Sinne von Inhaltsvalidität beim FWT von einer hohen Validität ausgegangen werden. Allerdings erlaubte die Eichstichprobe keine zusätzliche Erhebung externer kriteriumsbezogener Variablen, so dass eine Quantifizierung der Validität im Sinne einer Korrelation mit Außenkriterien nicht vorgenommen werden konnte.

Insgesamt stellt der FWT ein objektives, reliables und valides Messinstrument dar, welches einfach und ökonomisch in der Anwendung ist.

4. Itemanalyse

Zur Beurteilung der Güte einzelner Items können die Schwierigkeits- und Trennschärfekoeffizienten herangezogen werden. Der Schwierigkeitskoeffizient beurteilt das Verhältnis von falschen Lösungen zur Gesamtpersonenzahl, wobei ein Wert nahe 1 auf eine große Schwierigkeit des Items hinweist. Die Trennschärfe ist die Korrelation der Items mit der Gesamtskala. In der Tabelle finden sich sowohl die absoluten Trennschärfekoeffizienten (Korrelation mit Gesamtfragebogen) als auch die „corrected item-total correlations“ (Korrelation des i-ten Items mit der jeweils zugehörigen Skala ohne das i-te Item).

Die Daten in Tabellen 3 zeigen, dass die Trennschärfekoeffizienten der Items zwischen 0,33 und 0,66 liegen, mit einer starken Häufung zwischen 0,50 und 0,60. Jedes der Items indiziert damit den Gesamttestwert der Versuchsperson in zufriedenstellendem Maß.

Hinsichtlich der Schwierigkeit der Items ist aus Tabelle 3 zu ersehen, dass die Schwierigkeitsindizes bei der ersten Welle eine große Spannweite zwischen 0,28 und 0,97 aufweisen. Von der ersten zur zweiten Welle nimmt die Schwierigkeit dann zum Teil deutlich ab. Dies ist auf Lerneffekte zurückzuführen. Zwischen den ersten beiden Wellen wurde das Gebiet Wahrscheinlichkeitsrechnung im Durchschnitt 4 Wochen lang unterrichtet. Der Index des leichtesten Items (LW1) sinkt auf 0,08, das schwerste Item (GW1) liegt mit 0,85 immer noch recht hoch, so dass die Schwierigkeit der Items auch bei der zweiten Welle hinreichend streut. Bei der dritten Welle zeigen sich auf Itemebene nur geringe Unterschiede im Vergleich zur zweiten Welle.

Weiter zeigt sich, dass 5 von den 7 Itempaaren hinsichtlich Schwierigkeit und Trennschärfe als parallel angesehen werden können. Lediglich zwei Items (Laplace-Wahrscheinlichkeit und Kombinatorik) weisen systematisch verschiedene Lösungswahrscheinlichkeiten auf. Dies ist auf unterschiedliche Formulierungen der Items zurückzuführen (vgl. Anhang C). Allerdings führt gerade diese Unterschiedlichkeit der Schwierigkeiten bei der ersten Welle zu einer günstigeren Verteilung der Itemschwierigkeiten insgesamt, so dass diese Items unverändert in den Test übernommen wurden.

Teilgebiet	Item-name	Welle	Schwierigkeits-index	Trennschärfe (Testhälfte)	Trennschärfe (Gesamttest)
Bedingte Wahrscheinlichkeit	BW1	1	0,63	0,37	0,55
		2	0,44	0,35	0,57
		3	0,48	0,31	0,53
	BW2	1	0,70	0,27	0,51
		2	0,45	0,34	0,55
		3	0,50	0,29	0,52
Binomialverteilung	BI1	1	0,57	0,18	0,44
		2	0,37	0,16	0,45
		3	0,48	0,11	0,40
	BI2	1	0,56	0,14	0,42
		2	0,36	0,21	0,46
		3	0,46	0,10	0,37
Laplace-Wahrscheinlichkeit	LW1	1	0,28	0,44	0,62
		2	0,08	0,21	0,33
		3	0,09	0,19	0,33
	LW2	1	0,46	0,42	0,61
		2	0,24	0,29	0,42
		3	0,25	0,34	0,47
Erwartungswert	EW1	1	0,85	0,32	0,52
		2	0,56	0,34	0,59
		3	0,66	0,41	0,63
	EW2	1	0,81	0,32	0,51
		2	0,58	0,38	0,61
		3	0,61	0,38	0,62
Gegenwahrscheinlichkeit	GW1	1	0,97	0,26	0,37
		2	0,85	0,31	0,51
		3	0,89	0,36	0,52
	GW2	1	0,97	0,25	0,33
		2	0,83	0,36	0,55
		3	0,89	0,34	0,53
Kombinatorik	KO1	1	0,79	0,31	0,50
		2	0,55	0,24	0,43
		3	0,59	0,23	0,46
	KO2	1	0,59	0,33	0,52
		2	0,29	0,27	0,48
		3	0,40	0,31	0,51
Wahrscheinlichkeit stochastisch unabhängiger Ereignisse	UE1	1	0,66	0,47	0,66
		2	0,39	0,41	0,60
		3	0,38	0,38	0,61
	UE2	1	0,66	0,46	0,66
		2	0,37	0,43	0,60
		3	0,40	0,41	0,61

Tabelle 3: Schwierigkeiten und Trennschärfen der Items in den Wellen 1 - 3.

5. Veränderungsmessung mit dem FWT

Der FWT ist sowohl für die Messung des aktuellen Kenntnisstandes in WT als auch für die Messung der Veränderung des Kenntnisstandes, d. h. für die Messung von Lernzuwächsen, konzipiert. Veränderungsmessungen mit dem FWT erfolgt indirekt durch mehrfache Messungen im Längsschnitt⁴. Die Differenzenscores, also die Differenz der Testscores der Probanden zwischen zwei Untersuchungszeitpunkten, können als Maß für den individuellen Lernzuwachs interpretiert werden. Zwar wurde die Verwendung von Differenzenscores zur Erfassung intraindividuelle Veränderung früher oft kritisiert, jedoch können viele der Vorbehalte heute als überholt gelten (Rogosa, 1988). Auch werden manche Probleme durch die Modellierung von Veränderung mit latenten Differenzenscores gelöst (vgl. Steyer et al., 1997; Nachtigall et al., 2000). Der FWT bietet sich aufgrund seiner parallelen Testhälften für solche Formen der Veränderungsmodellierung an. Im Rahmen der Lernforschung können diese Differenzenscores, seien sie manifest berechnet oder latent im Rahmen von Strukturgleichungsmodellen ermittelt, auf mögliche erklärende Variablen zurückgeführt werden.

Die Interpretation der manifesten Differenzenscores bei mehrfachen Messungen mit FWT erfolgt in direkter Weise. Betrachtet man die Differenz zwischen späterer und früherer Welle, so geben die Differenzenscores die Veränderung im Kenntnisstand der Probanden bzgl. WT wieder. Positive Werte kennzeichnen Lernzuwächse, negative Werte kennzeichnen ein Nachlassen des Wissens in Wahrscheinlichkeitstheorie. Tabelle 4 liefert die Kennwerte der Verteilungen der Lernzuwächse; Prozenträge und Histogramme der Verteilungen finden sich in Anhang D.

Veränderung zwischen Welle	Stichproben-größe	Mittelwert	Standard-abweichung	Median
1 – 2	402	3,14	2,80	3,00
1 – 3	353	2,54	2,94	2,00
2 – 3	309	-0,64	2,60	-1,00

Tabelle 4: Mittelwerte und Standardabweichungen der Veränderung der Testwerte zwischen den Wellen 1-2, 1-3, 2-3.

Die Veränderung zwischen Welle 1 und 2 beschreibt dabei die Zeitspanne, während der WT in der Methodenausbildung des Psychologie-Grundstudiums unterrichtet wird. Im Mittel ist in dieser Zeitspanne ein Lernzuwachs von 3,14 Punkten zu beobachten, dies entspricht einer Effektstärke von $d=1,12$. Zwischen Welle 2 und 3 lagen die Semesterferien, in denen es im Mittel eine leichte Senkung der Kenntnisse gibt. Die Differenzen von Welle 1 und 3 geben die längerfristigen Lernzuwächse von Beginn der Behandlung des Stoffes im ersten Semester bis hin zu Beginn des zweiten Semesters wieder.

⁴ Zur Diskussion über direkte und indirekte Veränderungsmessung vgl. z.B. Petermann (1978).

Literatur

- Feuerpfeil, J. & Heigl, F. (1991). Wahrscheinlichkeitsrechnung und Statistik. Bayerischer Schulbuch-Verlag München. 3. Auflage.
- Griesel, H. & Postel, H. (Hrsg.) (1990). Grundkurs Stochastik. Hannover: Schroedel Schulbuchverlag.
- Nachtigall, C., Kämpfe, N., Kärtner, J., Kraus, K., Preißler, S. & Wolf, A. 2001. Die Jenaer Panel-Studie zu Lernzuwächsen im Fach Wahrscheinlichkeitstheorie. metheval report 3, (3). <http://www.uni-jena.de/svw/metheval/report/>.
- Nachtigall, C., Kraus, K. & Steyer, R. (2000). The analysis of change: True change models and growth curves. Proceedings of the 5th International Conference on Social Science Methodology, Cologne. <http://www.uni-jena.de/svw/metheval/publikationen/>.
- Petermann, F. (1978). Veränderungsmessung. Stuttgart: Kohlhammer.
- Plachky, D., Baringhaus, L. & Schmitz, N. (1977). Stochastik I. Eine elementare Einführung in Grundbegriffe der Wahrscheinlichkeitstheorie und Statistik. Akademische Verlagsgesellschaft Wiesbaden.
- Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), Methodological issues in aging research (pp. 171 - 219).
- Schmid, A. & Schweizer, W. (Hrsg.) (1988). Stochastik Leistungskurs. Stuttgart: Klett.
- Schulmeister, R. (Hrsg.) (1983). Angst vor Statistik. Hamburg: Arbeitsgemeinschaft für Hochschuldidaktik. Heft 87.
- Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. Methods of Psychological Research-Online, 2, 21-33.
- Steyer, R. & Eid, M. (1994). Messen und Testen. Berlin: Springer.
- Wolf, A. & Nachtigall, C. (2001). Computerversion des Fragebogens zur Wahrscheinlichkeitstheorie (FWT). metheval report, in Druck, <http://www.uni-jena.de/svw/metheval/report/>.

Anhang

A. Prozentränge und Verteilungen der Testscores

Perzentil	Welle 1	Welle 2	Welle 3
10	0,30	3,00	3,00
20	2,00	5,00	4,00
30	2,00	6,00	5,00
40	3,00	7,00	6,00
50	4,00	8,00	7,00
60	5,00	9,00	8,00
70	6,00	10,00	9,00
80	7,00	11,00	10,00
90	9,00	12,00	12,00
100	14,00	14,00	14,00

Tabelle 5: Prozentränge der Testwerte in den Wellen 1 – 3.

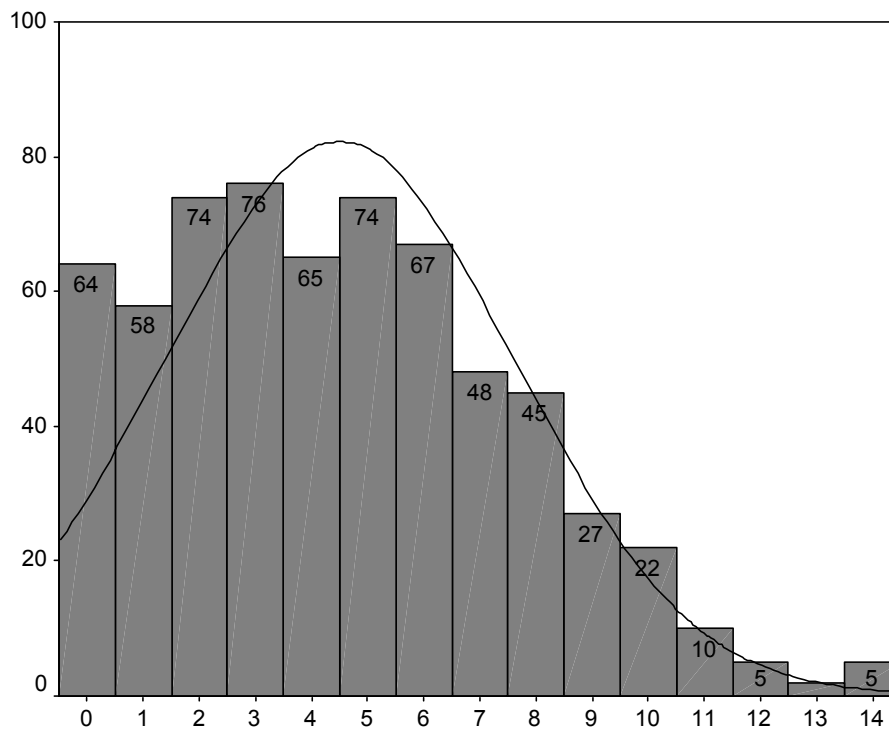


Abbildung 1: Verteilung der Testwerte in Welle 1.

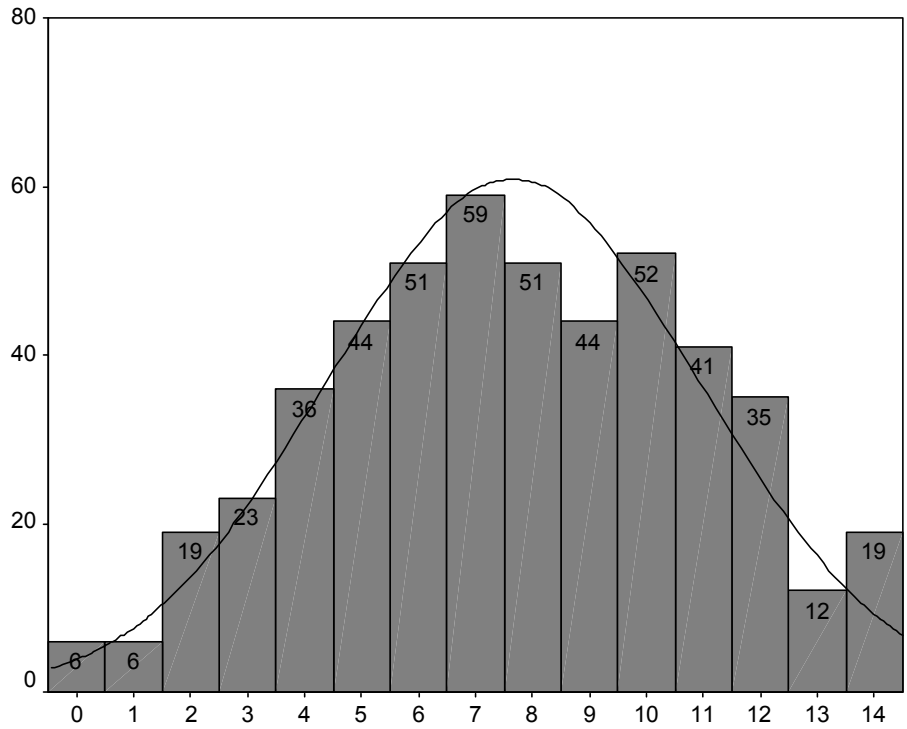


Abbildung 2: Verteilung der Testwerte in Welle 2.

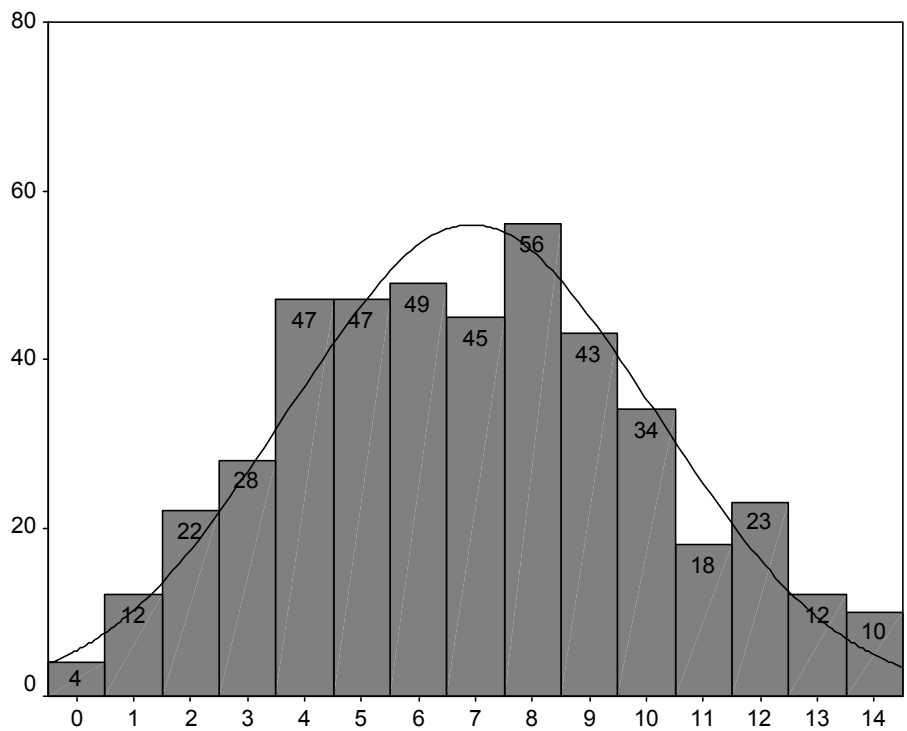


Abbildung 3: Verteilung der Testwerte in Welle 3.

Die Verteilung kann angenähert durch eine Normalverteilung beschrieben werden, wobei sich bei der ersten Welle leichte Bodeneffekte zeigen. Die Deckeneffekte bei Welle 2 und 3 sind dagegen als gering zu bewerten.

B. Korrelationen aller Items zu jedem der 3 Messzeitpunkte

	1. Welle	2. Welle	3. Welle
1. Welle	9,65	0,61	0,55
2. Welle	6,04	10,62	0,67
3. Welle	5,27	6,81	10,25

Tabelle 6: Kovarianz- und Korrelationsmatrix der Gesamtpunkte der 3 Wellen. Die Kovarianzen befinden sich in der linken unteren Dreiecksmatrix, die Korrelation (fettgedruckte Werte) rechts über der Hauptdiagonalen.

	BW 1	BW 2	BI 1	BI 2	LW 1	LW 2	EW 1	EW 2	GW 1	GW 2	KO 1	KO 2	UE 1	UE 2
BW 1														
BW 2	0,53													
BI 1	0,10	0,06												
BI 2	0,08	0,02	0,78											
LW 1	0,35	0,32	0,16	0,11										
LW 2	0,31	0,26	0,09	0,08	0,59									
EW 1	0,24	0,26	0,08	0,09	0,21	0,20								
EW 2	0,20	0,21	0,07	0,08	0,21	0,24	0,73							
GW 1	0,11	0,13	0,13	0,09	0,11	0,16	0,22	0,21						
GW 2	0,08	0,12	0,13	0,07	0,10	0,13	0,20	0,17	0,81					
KO 1	0,17	0,19	0,08	0,07	0,21	0,22	0,17	0,15	0,18	0,16				
KO 2	0,14	0,11	0,12	0,14	0,21	0,24	0,16	0,15	0,16	0,16	0,47			
UE 1	0,28	0,18	0,14	0,12	0,37	0,38	0,26	0,27	0,21	0,17	0,31	0,30		
UE 2	0,25	0,21	0,13	0,12	0,38	0,39	0,28	0,25	0,21	0,21	0,30	0,32	0,68	

Tabelle 7: Korrelationsmatrix der einzelnen Items in Welle 1.

	BW 1	BW 2	BI 1	BI 2	LW 1	LW 2	EW 1	EW 2	GW 1	GW 2	KO 1	KO 2	UE 1	UE 2
BW 1														
BW 2	0,59													
BI 1	0,09	0,11												
BI 2	0,08	0,08	0,84											
LW 1	0,16	0,17	0,11	0,13										
LW 2	0,18	0,17	0,12	0,13	0,33									
EW 1	0,27	0,20	0,12	0,16	0,11	0,19								
EW 2	0,29	0,23	0,11	0,15	0,07	0,20	0,79							
GW 1	0,17	0,18	0,16	0,16	0,10	0,14	0,17	0,18						
GW 2	0,20	0,21	0,19	0,19	0,11	0,17	0,19	0,19	0,88					
KO 1	0,17	0,14	0,05	0,03	0,01	0,05	0,13	0,17	0,16	0,17				
KO 2	0,22	0,17	0,06	0,02	0,08	0,10	0,22	0,24	0,18	0,20	0,43			
UE 1	0,26	0,30	0,07	0,09	0,23	0,19	0,28	0,30	0,23	0,26	0,24	0,27		
UE 2	0,28	0,29	0,15	0,16	0,18	0,22	0,25	0,27	0,20	0,25	0,21	0,21	0,56	

Tabelle 8: Korrelationsmatrix der einzelnen Items in Welle 2.

	BW 1	BW 2	BI 1	BI 2	LW 1	LW 2	EW 1	EW 2	GW 1	GW 2	KO 1	KO 2	UE 1	UE 2
BW 1														
BW 2	0,58													
BI 1	0,08	0,04												
BI 2	0,06	0,02	0,84											
LW 1	0,08	0,12	0,00	0,00										
LW 2	0,14	0,18	0,06	0,07	0,37									
EW 1	0,29	0,29	0,09	0,10	0,16	0,14								
EW 2	0,26	0,25	0,08	0,07	0,17	0,16	0,77							
GW 1	0,19	0,16	0,16	0,14	0,08	0,13	0,30	0,24						
GW 2	0,20	0,17	0,17	0,15	0,08	0,15	0,31	0,26	0,93					
KO 1	0,11	0,13	0,06	0,00	0,08	0,22	0,14	0,17	0,23	0,21				
KO 2	0,13	0,15	0,11	0,05	0,12	0,24	0,17	0,16	0,19	0,20	0,51			
UE 1	0,25	0,26	0,03	0,00	0,24	0,25	0,35	0,37	0,18	0,19	0,17	0,25		
UE 2	0,25	0,21	0,02	0,03	0,24	0,30	0,36	0,36	0,19	0,19	0,17	0,26	0,70	

Tabelle 9: Korrelationsmatrix der einzelnen Items in Welle 3.

C. Die Items des FWT

- BW 1 Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf eine gerade Augenzahl geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl erbracht hat, die größer als drei ist?
- BW 2 Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf zwei oder weniger Augen geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl kleiner als fünf erbracht hat?
- BI 1 Wie wahrscheinlich ist es, dass Sie bei zwei Würfeln mit einer fairen 10-Pfennig-Münze genau einmal "Eiche" werfen?
- BI 2 Wie wahrscheinlich ist es, dass Sie bei zwei Würfeln mit einem fairen 5-DM-Stück genau einmal "Zahl" werfen?
- LW 1 Wie wahrscheinlich ist es, bei einem Würfelwurf eine Augenzahl zu werfen, die höchstens zwei ist?
- LW 2 Wie wahrscheinlich ist es, bei einem Würfelwurf eine Augenzahl zu werfen, die mindestens drei ist?
- EW 1 Wie viele Augen sind durchschnittlich insgesamt zu erwarten, wenn Sie dreimal würfeln und die Augenzahlen zusammenzählen?
- EW 2 Wie viele Felder erwarten Sie bei einem Brettspiel im Durchschnitt vorrücken zu können, wenn Sie fünfmal hintereinander würfeln dürfen?
- GW 1 Wie wahrscheinlich ist es, dass Sie bei viermaligem Würfeln mindestens eine eins werfen? (Ausrechnen nicht nötig, Formel angeben genügt.)
- GW 2 Wie wahrscheinlich ist es, dass Sie bei fünfmaligem Würfeln mindestens eine sechs werfen? (Ausrechnen nicht nötig, Formel angeben!)
- KO 1 Wie viele verschiedene Möglichkeiten gibt es, vier verschiedene Süßigkeiten an vier verschiedene Kinder zu verteilen, wenn jedes Kind nur eine Süßigkeit bekommt?
- KO 2 Drei Personen steigen nacheinander aus dem Bus aus. Wie viele verschiedene Möglichkeiten gibt es dabei?
- UE 1 Wie wahrscheinlich ist es, bei dreimaligem Würfeln im ersten Wurf eine vier, im zweiten Wurf eine sechs und im dritten Wurf eine eins zu werfen?
- UE 2 Wie wahrscheinlich ist es, bei drei Münzwürfen mit einer 10-Pfennig-Münze beim ersten Wurf "Eiche", beim zweiten Wurf "Zahl" und beim dritten Wurf "Eiche" zu werfen?

Die Aufgaben wurden in Form A in folgender Reihenfolge im Test abgefragt: LW2, LW1, BW2, BW1, EW2, EW1, UE1, UE2, BI1, BI2, GW2, GW1, KO1, KO2.

Bei Form B des Fragebogens lautet die Reihenfolge:

LW1, LW2, BI2, BI1, GW2, GW1, KO2, KO1, BW1, BW2, UE2, UE1, EW1, EW2.

D. Prozentränge und Verteilungen der Differenzenscores

Perzentil	Welle 1 - 2	Welle 1 - 3	Welle 2 - 3
10	,00	-1,00	-4,00
20	1,00	,00	-3,00
30	2,00	1,00	-2,00
40	2,00	2,00	-1,00
50	3,00	2,00	-1,00
60	4,00	3,00	,00
70	4,00	4,00	1,00
80	5,00	5,00	2,00
90	7,00	6,00	2,00
100	12,00	12,00	10,00

Tabelle 10: Prozentränge der Veränderung der Testwerte zwischen den Wellen 1-2, 1-3, 2-3.

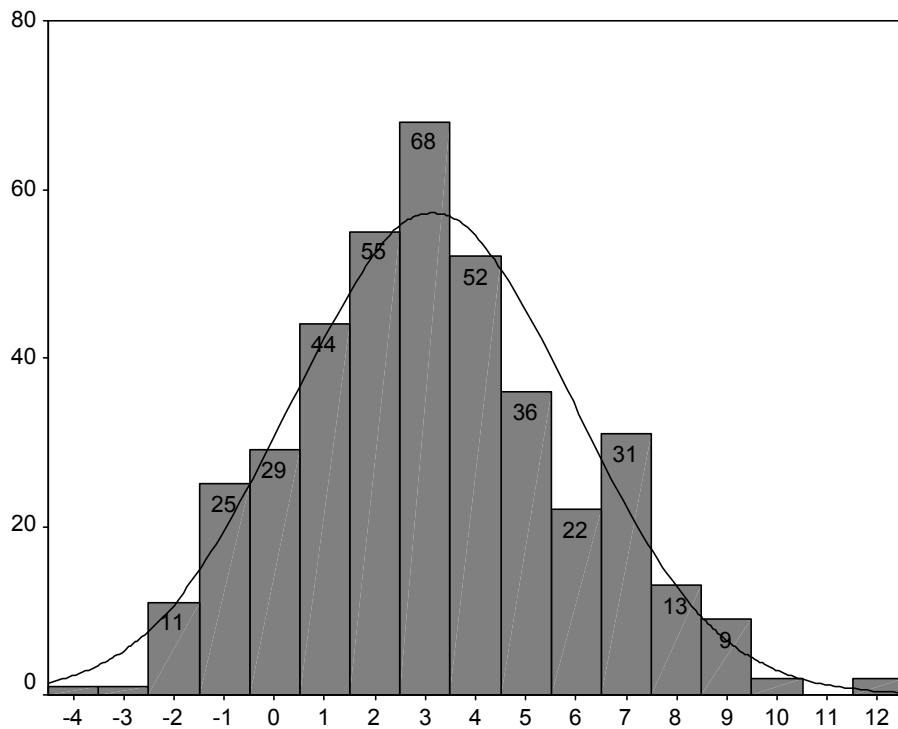


Abbildung 4: Verteilung der Differenzen zwischen den Testwerten der Wellen 1 und 2.

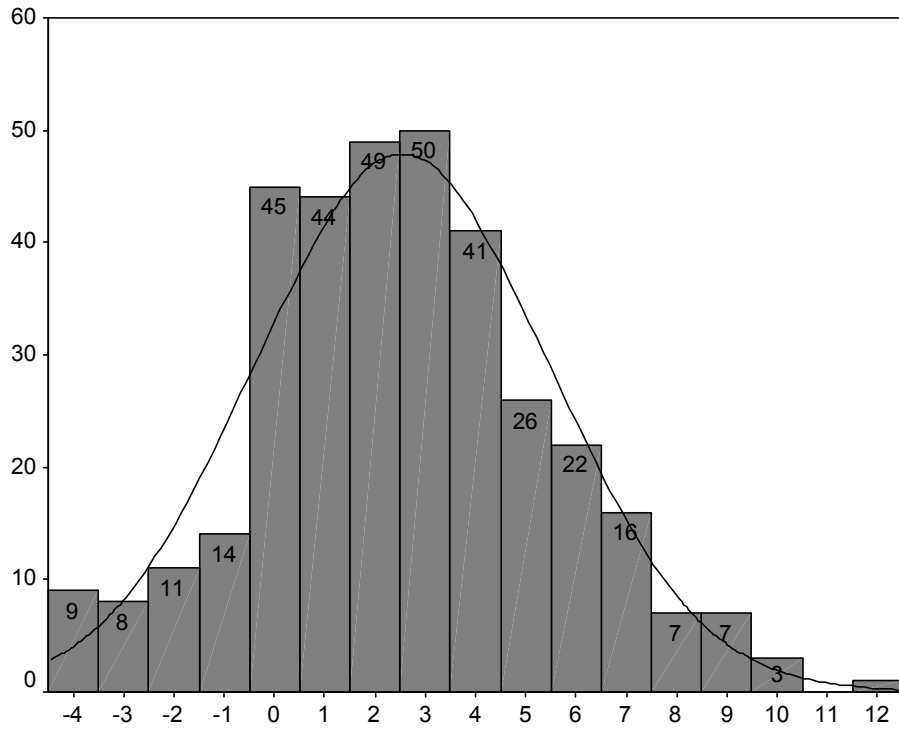


Abbildung 5: Verteilung der Differenzen zwischen den Testwerten der Wellen 1 und 3.

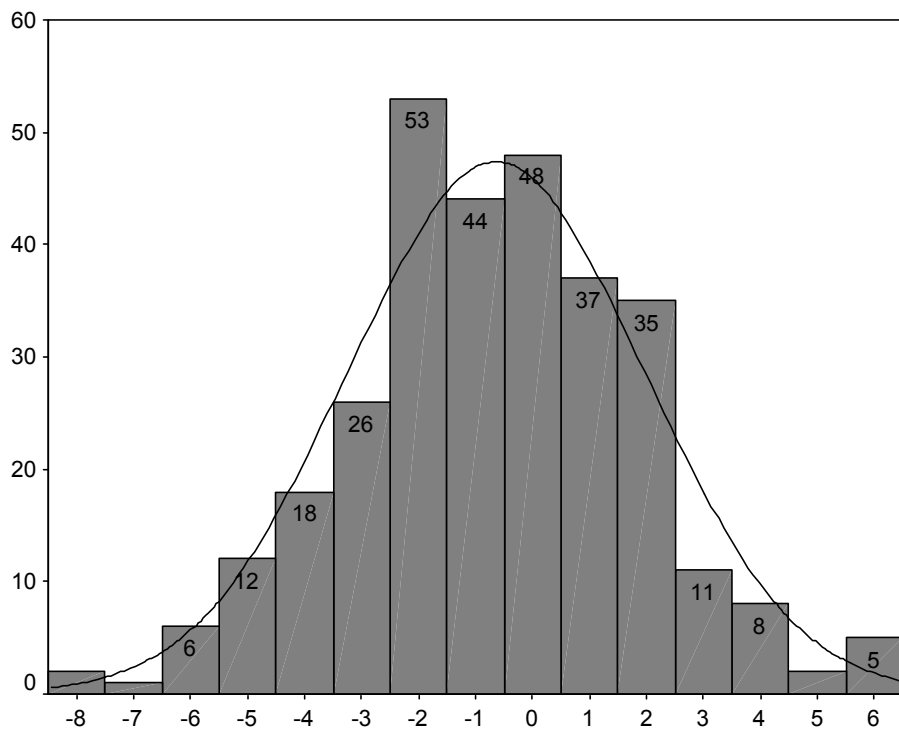


Abbildung 6: Verteilung der Differenzen zwischen den Testwerten der Wellen 2 und 3.