

# Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität

Jochen Prümper

## Zusammenfassung

In diesem Beitrag wird die testtheoretische Güte des Benutzungsfragebogens ISONORM 9241/10 [17] dokumentiert. Grundlage hierzu sind Beurteilungen von 1265 Benutzern von insgesamt 178 Softwareprogrammen. Zur Bestimmung der testtheoretischen Güte werden Untersuchungen zur Reliabilität, Validität und Normierung durchgeführt. Zur Reliabilitätsbestimmung wird die *Wiederholungsmethode* und die *Konsistenzanalyse* herangezogen. Die *Wiederholungsmethode* zeigt, daß sich der ISONORM 9241/10 durch eine hohe Merkmalskonstanz auszeichnet. Die *Konsistenzanalyse* zeigt die Leistungsfähigkeit des ISONORM 9241/10 als Meßinstrument, da die einzelnen Skalen des ISONORM 9241/10 gute innere Konsistenzen aufweisen. Im Rahmen der Validitätsbestimmung werden Untersuchungen zur *kriterienbezogenen Validität* und zur *Konstruktvalidität* angestellt. Zur Überprüfung der *inneren kriterienbezogenen Validität* werden die ISONORM 9241/10-Bewertungen mit den Bewertungen von zwei Benutzungsfragebögen QUIS [21] und BBD [22] korreliert. Es zeigt sich, daß der ISONORM 9241/10 zu vergleichbaren Ergebnissen gelangt, und gleichzeitig den Vorteil in sich birgt, eine ausdrückliche Operationalisierung von ISO 9241 Teil 10 darzustellen. Zur Überprüfung der *äußeren kriterienbezogenen Validität* wird unter Verwendung von EVADIS II [9] überprüft, inwiefern die Befragung durch Benutzer mit der Evaluation durch Experten korrespondiert. Es zeigt sich, daß die ISONORM 9241/10-Benutzungsanalysen einen signifikanten Bezug zu einer systematischen, leitfadensorientierten Expertenevaluation aufweisen. Zur Überprüfung der *Konstruktvalidität* werden zwei Versionen eines Softwareprogramms mit und zwei Versionen eines Softwareprogramms ohne graphische Benutzungsschnittstelle herangezogen. Softwareergonomische Veränderungen von einer Version zur nächsten sind mittels ISONORM 9241/10 meßbar; zusätzlich lassen sich differenzierte Aussagen über diese Verbesserungen anstellen. Die *Normierung* erlaubt die Einordnung einer einzelnen Software in Relation zu den Beurteilungen anderer Programme.

## 1 Einleitung

Der Benutzungsfragebogen ISONORM 9241/10 wurde erstmals auf der Software-Ergonomie Fachtagung '93 vorgestellt [17]. Er stellt eine Operationalisierung der ISO 9241 Teil 10 „Grundsätze der Dialoggestaltung“ dar, welchem eine zentrale Rolle im Rahmen der „Ergonomischen Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ [4] zukommt. Die dort formulierten Forderungen sind unabhängig von einer bestimmten Dialogtechnik und sollen sowohl bei der Leistungsbeschreibung, der Gestaltung als auch der Bewertung von Dialogsystemen angewandt werden. Erklärtes Ziel der Anwendung dieser Norm ist es, Benutzungsschnittstellen zu gewährleisten, „die gebrauchstauglicher und konsistenter sind und eine höhere Produktivität ermöglichen“ [4, S. 4]. Zu diesem Zweck werden in ihr sieben allgemeine ergonomische Grundsätze definiert, die mit Empfehlungen und Beispielen konkretisiert und erläutert werden. Bei diesen Grundsätzen handelt es sich um: Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit und Lernförderlichkeit.

Bei der Entwicklung des ISONORM 9241/10 stand im Vordergrund, ein effizient einsetzbares, robustes Verfahren zu entwickeln, das zur Überprüfung der Normkonformität von ISO

9241/10 herangezogen werden kann. Ein weiteres Anwendungsgebiet ist die Beurteilung von Prototypen im Rahmen eines iterativen Systemdesigns.

Erste Untersuchungen wurden hierzu bereits vorgelegt [12,14,15,16,17]. Es wurde dargestellt, daß der ISONORM 9241/10 ein nützliches formatives Evaluationsverfahren ist, wenn er im Rahmen eines partizipativen Designkonzeptes dazu genutzt wird, in Kombination mit einer visualisierten Gruppendiskussion erste konkrete Schwachstellen einer Software zu identifizieren [17].

In einer summativen Evaluationsstudie über die ergonomische Qualität von Softwareprodukten mit und ohne graphische Benutzungsschnittstellen (GUI) konnte - im Sinne einer ersten Validierung - erwartungsgemäß gezeigt werden, daß Software mit GUI auf allen sieben Skalen des ISONORM 9241/10 signifikant bessere Werte erreichte als solche ohne GUI [15].

Dennoch weisen die Autoren des ISONORM 9241/10 selbst darauf hin, daß zur Überprüfung der testtheoretischen Qualität des Verfahrens - insbesondere hinsichtlich der Reliabilität und Validität - noch weitere Untersuchungen angestellt werden müssen [15,17]. Dieses Defizit zu beheben ist Thema des vorliegenden Beitrags. Hierzu erfolgt zunächst eine kurze Beschreibung des ISONORM 9241/10 sowie der Benutzer, die an den Evaluationen teilgenommen haben. Anschließend wird der Nachweis der Reliabilität anhand der beiden Verfahren *Wiederholungsmethode* und *Konsistenzanalyse* geleistet. Zur Validitätsbestimmung werden Untersuchungen zur *kriterienbezogenen Validität* und zur *Konstruktvalidität* angestellt.

### 1.1 Beschreibung des Verfahrens

Um dem Anspruch eines ökonomisch einsetzbaren Verfahrens gerecht zu werden, wurde der Umfang des ISONORM 9241/10-Fragebogens so zugeschnitten, daß jeder einzelne der sieben Grundsätze von ISO 9241/10 über fünf Items operationalisiert wurde. Insgesamt umfaßt das Verfahren damit 35 Items.

Der „Granulationsgrad“ [19] der Beschreibungssprache wurde so gewählt, daß die spezifischen Eigenschaften unterschiedlicher Softwaresysteme hinreichend genau differenziert werden können und er gleichzeitig auf möglichst viele unterschiedliche Programme anwendbar ist. Die Durchführungszeit für den ISONORM 9241/10 beträgt ca. 10 Minuten.

Jeder der sieben Grundsätze von ISO 9241/10 wurde über fünf bipolare Aussagen konkretisiert. Beantwortet werden diese Aussagen auf einer sieben-stufigen Skala von „- - -“ (1) bis „+ + +“ (7). Abbildung 1 liefert ein Beispielitem des ISONORM 9241/10 für den Grundsatz „Steuerbarkeit“.

Die Software ...	---	--	-	-/+	+	++	+++	Die Software ...
erzwingt eine unnötig starre Einhaltung von Bearbeitungsschritten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	erzwingt keine unnötig starre Einhaltung von Bearbeitungsschritten.

Abb. 1: Beispielitem des ISONORM 9241/10 für den Grundsatz „Steuerbarkeit“

Den Benutzungsfragebogen ISONORM 9241/10 gibt es in deutscher, englischer, französischer und niederländischer Sprache.

## 1.2 Beschreibung der Gesamtstichprobe

Bislang liegen von 1265 Benutzern (51,8% Frauen, 48,2% Männer) Beurteilungen von insgesamt 178 Softwareprogrammen vor. Das Durchschnittsalter der Befragten lag bei rund 34 Jahren, sie arbeiteten im Durchschnitt seit 77 Monaten mit Computern und seit 25 Monaten mit der von ihnen beurteilten Software. Die Frage „Wie gut beherrschen Sie die beurteilte Software?“ beantworteten die befragten Personen auf einer Skala von 1 (sehr schlecht) bis 7 (sehr gut) im Durchschnitt mit 5,1. Mit der beurteilten Software arbeiteten die Personen im Mittel 14 Stunden pro Woche, mit Computern überhaupt 24 Stunden pro Woche. Insgesamt arbeiteten die Befragten durchschnittlich mit 5,8 verschiedenen Softwareprogrammen.

## 2 Reliabilität

Unter der *Reliabilität* oder Zuverlässigkeit eines Meßverfahrens versteht man den Grad der Genauigkeit, mit dem es ein bestimmtes Merkmal mißt, gleichgültig, ob es dieses Merkmal auch zu messen beansprucht (letzteres ist eine Frage der Validität) [7]. Zwei Vorgehensweisen der Reliabilitätsbestimmung sollen im folgenden beleuchtet werden: die *Wiederholungsmethode* und die *Konsistenzanalyse*.

### 2.1 Wiederholungsmethode

Die *Wiederholungsmethode* - oder auch Bestimmung der *Retest-Reliabilität* - schlägt sich in einem Stabilitätsmaß nieder und gibt an, inwiefern eine Person bei Befragung zu einem späteren Zeitpunkt zu ähnlichen Beurteilungen desselben Sachverhaltes gelangt.

Zur Bestimmung der Retest-Reliabilität wurde eine Substichprobe von 49 Benutzern gebeten, ihre ISONORM 9241/10-Beurteilung zu einem zweiten Zeitpunkt für dieselbe Software zu wiederholen. Zwischen den Meßzeitpunkten  $t_1$  und  $t_2$  lagen im Mittel 6,7 Monate.

Tabelle 1 gibt die Retest-Reliabilitäten des ISONORM 9241/10 wieder. Die niedrigste Retest-Reliabilität beträgt bei dem Grundsatz Lernförderlichkeit  $r_{t1/t2} = .59$ , die höchste Retest-Reliabilität bei dem Grundsatz Fehlertoleranz  $r_{t1/t2} = .68$ . Die durchweg hochsignifikanten Stabilitätskoeffizienten zeigen, daß Benutzer auch nach einem längeren Zeitraum zu ähnlichen Beurteilungen desselben Softwaresystems gelangen. Der ISONORM 9241/10 legt damit einen zufriedenstellenden Grad an Stabilität an den Tag.

### 2.2 Konsistenzanalyse

Die *Konsistenzanalyse* untersucht als Methode zur Bestimmung der Skalenzuverlässigkeit den Grad der Homogenität der einzelnen Items im Hinblick auf das zu erfassende Merkmal. Diese innere Konsistenz kennzeichnet als instrumentelle Reliabilität gewissermaßen die Leistungsfähigkeit des Meßinstrumentes und ist ein von den Bedingungen des Meßdurchführung unabhängiger Aspekt der Reliabilität [7].

Zur Bestimmung der *inneren Konsistenz* konnte wiederum die Gesamtstichprobe analysiert werden. Die Berechnung der inneren Konsistenz erfolgt über Cronbachs alpha.

Es zeigt sich (vgl. Tab. 1), daß die einzelnen Skalen des ISONORM 9241/10 gute innere Konsistenzen aufweisen.

Grundsatz	$r_{1/2}^*$	$\alpha^{\circ}$
Aufgabenangemessenheit	.67	.81
Selbstbeschreibungsfähigkeit	.62	.86
Steuerbarkeit	.64	.84
Erwartungskonformität	.60	.84
Fehlertoleranz	.68	.87
Individualisierbarkeit	.63	.89
Lernförderlichkeit	.59	.83

Anmerkung: \*  $p < .001$ ;  $N = 49$  Benutzer  
 $\circ$   $N$  zwischen 1208 und 1251 Benutzern;  
Bestimmung der inneren Konsistenz über  
Cronbachs alpha

Tab. 1: Retest-Reliabilität und innere Konsistenz des ISONORM 9241/10

### 3 Validität

Die *Validität* oder Gültigkeit eines Tests gibt den Grad der Genauigkeit an, mit dem ein Meßverfahren das, was es zu messen beansprucht, tatsächlich mißt [7]. Bei der Validität sollen verschiedene Aspekte unterschieden werden: zum einen die *kriterienbezogene Validität* und zum anderen die *Konstruktvalidität*.

#### 3.1 Kriterienbezogene Validität

Bei der *kriterienbezogenen Validität* werden die Meßergebnisse mit einem sog. Außenkriterium in Verbindung gesetzt. Bezüglich der kriterienbezogenen Validität läßt sich eine Unterscheidung zwischen innerer und äußerer Validität fällen. Von *innerer Validität* spricht man, wenn der Skalenwert mit den Skalenwerten anderer schon als valide anerkannter Meßverfahren korreliert wird. Unter *äußerer Validität* versteht man das Korrelieren der Skalenwerte mit einem Schätzurteil oder mit einer objektiv bewerteten Kriteriumsleistung [7].

##### 3.1.1 Innere kriterienbezogene Validität

Zur Überprüfung der inneren kriterienbezogenen Validität soll der Frage nachgegangen werden, ob der ISONORM 9241/10 zu vergleichbaren Ergebnissen gelangt wie andere Benutzungsfragebögen.

Für diese Validierung wurden zwei Fragebögen herangezogen, die ebenfalls Benutzungsfreundlichkeit von Software erheben. Es handelt sich dabei um die deutsche Übersetzung [6] der Langform des QUIS [21] und um den BBD [22]. Da beide Verfahren keinen Bezug zu ISO 9241/10 aufweisen, muß sich die Validitätsuntersuchung auf den Vergleich der jeweiligen Gesamtwerte beschränken.

Für diese Studie wurden 31 Benutzer gebeten, eine ihnen bekannte Software sowohl anhand des ISONORM 9241/10 als auch mittels des QUIS und des BBD zu beurteilen. Diese Benutzer arbeiteten durchschnittlich seit rund 18 Monaten mit der von ihnen evaluierten Software und seit rund 58 Monaten überhaupt mit Computern. Die Frage „Wie gut beherrschen Sie die beurteilte Software?“ beantworteten sie auf einer Skala von 1 (sehr schlecht) bis 7 (sehr gut) im Durchschnitt mit 5.6.

Die Korrelation zwischen ISONORM 9241/10 und QUIS beträgt  $r = .73$  ( $p < .001$ ;  $N = 31$  Benutzer), die zwischen ISO 9241/10 und BBD  $r = .71$  ( $p < .001$ ;  $N = 31$  Benutzer).

Diese Korrelationen zeigen, daß der ISONORM 9241/10 in der Gesamtbetrachtung zu ähnlichen Ergebnissen gelangt wie andere Benutzungsfragebögen, jedoch den Vorteil in sich birgt, eine explizite Operationalisierung von ISO 9241 Teil 10 zu sein.

### 3.1.2 Äußere kriterienbezogene Validität

Zur Bestimmung der äußeren kriterienbezogenen Validität soll überprüft werden, ob die Befragung durch Benutzer mit der Evaluation durch Experten korrespondiert.

EVADIS II [9] ist ein solches softwareunterstütztes, leitfadenorientiertes Evaluationsverfahren für Ergonomieexperten.

Für diese Validitätsstudie wurden 13 verschiedene Softwareprogramme von einem Software-Ergonomen in Zusammenarbeit mit einem erfahrenen Benutzer der jeweiligen Software anhand von EVADIS II evaluiert und mit den entsprechenden ISONORM 9241/10-Beurteilungen durch Benutzer verglichen. Für die ISONORM 9241/10-Analysen wurde eine Substichprobe von 383 Benutzern herangezogen. Jedes der 13 Softwareprogramme wurde dabei durchschnittlich von 29.5 ( $N_{\min} = 11$ ,  $N_{\max} = 79$ ) Benutzern beurteilt.

Die Korrelation zwischen dem EVADIS II-Gesamtwert und dem ISONORM 9241/10-Gesamtwert beträgt  $r = .59$  ( $p < .01$ ;  $N = 13$  Softwareprogramme).

Es zeigt sich also, daß die ISONORM 9241/10-Benutzeranalysen einen signifikanten Bezug zu einer systematischen, leitfadenorientierten Expertenevaluation aufweisen.

## 3.2 Konstruktvalidität

Bei der *Konstruktvalidität* leitet man Hypothesen aus dem zugrundeliegenden Konstrukt ab und überprüft diese empirisch mittels des zu validierenden Meßinstrumentes [7].

Im vorliegenden Fall liegt der Bestimmung der Konstruktvalidität die Hypothese zugrunde, daß bei der neuen Version einer Software ergonomische Qualitätsverbesserungen zu verzeichnen sind. Im Sinne einer Validierung sollten diese Qualitätsverbesserungen mittels ISONORM 9241/10 meßbar sein. Dabei soll insbesondere ein Augenmerk darauf gelegt werden, ob sich für verschiedene Softwareprogramme differenzierte Aussagen hinsichtlich der Qualitätsverbesserung bei einzelnen Grundsätzen anstellen lassen.

Für diese Validitätsstudie sollten zwei Versionen eines Softwareprogramms mit und zwei Versionen eines Softwareprogramms ohne graphische Benutzungsschnittstelle herangezogen werden. Exemplarisch für diese Fragestellung soll MS-Word für DOS (mit den Versionen 4.0 und 5.0) sowie MS-Excel für Windows (mit den Versionen 4.0 und 5.0) einer näheren Betrachtung unterzogen werden.

Eine Substichprobe von 118 Benutzern wurde für diese ISONORM 9241/10-Analysen herangezogen. MS-Word 4.0 wurde dabei von 22 Benutzern, MS-Word 5.0 von 52 Benutzern, MS-Excel 4.0 von 28 Benutzern und MS-Excel 5.0 von 16 Benutzern evaluiert. Zwischen den entsprechenden Gruppen gab es keine signifikanten Unterschiede hinsichtlich der erhobenen Expertise- und Demographie-Daten.

Abbildung 2 verdeutlicht die Ergebnisse für den Versionsvergleich von MS-Word 4.0 zu 5.0 und Abbildung 3 den Versionsvergleich von MS-Excel 4.0 zu 5.0 für die einzelnen Skalen des ISONORM 9241/10.

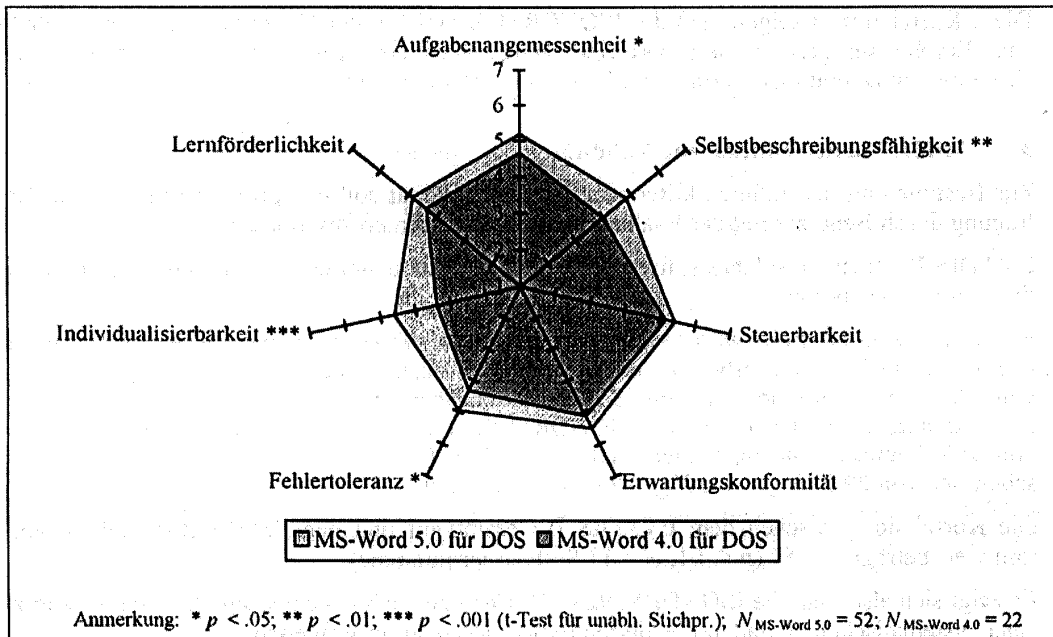


Abb. 2: ISONORM 9241/10 - Versionsvergleich von MS-Word 4.0 und MS-Word 5.0

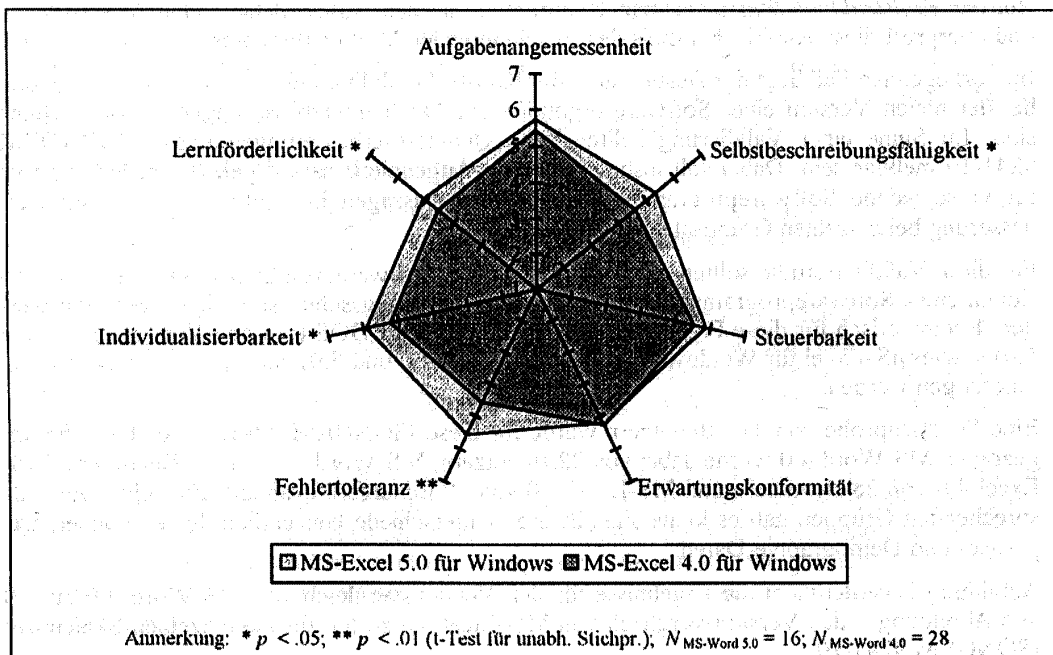


Abb. 3: ISONORM 9241/10 - Versionsvergleich von MS-Excel 4.0 und MS-Excel 5.0

Es zeigt sich, daß sich MS-Word 5.0 gegenüber MS-Word 4.0 hinsichtlich der vier Grundsätze Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Fehlertoleranz und Individualisierbarkeit signifikant verbessert hat. Dabei unterscheidet sich der ISONORM 9241/10-Gesamtwert signifikant auf dem 5%-Niveau mit folgenden Mittelwerten:  $\bar{x}_{\text{MS-Word 5.0}} = 5.06$  und  $\bar{x}_{\text{MS-Word 4.0}} = 4.45$ .

Bei MS-Excel sieht das Bild anders aus: MS-Excel 5.0 verbesserte sich gegenüber MS-Excel 4.0 signifikant in puncto Selbstbeschreibungsfähigkeit, Fehlertoleranz, Individualisierbarkeit und Lernförderlichkeit. Der ISONORM 9241/10-Gesamtwert unterscheidet sich dabei signifikant auf dem 5%-Niveau wie folgt:  $\bar{x}_{\text{MS-Excel 5.0}} = 5.52$  und  $\bar{x}_{\text{MS-Excel 4.0}} = 5.03$ .

Damit zeigt sich, daß Veränderungen von einer Version zur nächsten mittels ISONORM 9241/10 meßbar sind, und daß sich zusätzlich differenzierte Aussagen über diese Verbesserungen anstellen lassen.

#### 4 Normierung des ISONORM 9241/10

Unter der Normierung eines Meßinstrumentes versteht man das Erstellen eines Bezugssystems, das die Einordnung eines individuellen Meßergebnisses ermöglicht. Ziel ist es, die Vergleichbarkeit einzelner Messungen zu gewährleisten [7].

Im vorliegenden Fall soll diese Normierung aus der Ermittlung der Verteilung der Meßwerte und bestimmter Verteilungsparameter bestehen. Dadurch besteht die Möglichkeit, den ISONORM 9241/10-Wert einer einzelnen Software in Relation zu den Beurteilungen anderer Programme zu setzen.

Grundlage dieser Normierung sollten nur die Softwareprogramme sein, die von mindestens sieben Benutzern beurteilt wurden. Dieses Kriterium erfüllten 41 Softwareprogramme. Jedes dieser 41 Softwareprogramme wurde dabei durchschnittlich von 24.6 ( $N_{\text{min}} = 7$ ,  $N_{\text{max}} = 106$ ) Benutzern beurteilt.

Tabelle 2 gibt die Skalenmittelwerte, das jeweilige Minimum und Maximum sowie die Standardabweichungen an. Alle sieben Skalen sind normalverteilt.

Grundsatz	$\bar{x}$	min.	max.	s
Aufgabenangemessenheit	4.87	2.60	5.76	.70
Selbstbeschreibungsfähigkeit	4.55	2.50	5.70	.65
Steuerbarkeit	5.01	3.20	6.13	.75
Erwartungskonformität	4.94	2.97	6.09	.72
Fehlertoleranz	4.29	2.59	5.75	.79
Individualisierbarkeit	4.28	2.41	5.92	.92
Lernförderlichkeit	4.52	2.56	5.83	.70

Anmerkung:  $N = 41$  Programme

Tab. 2: Skalenwerte des ISONORM 9241/10

Der niedrigste ISONORM 9241/10-Wert eines Softwaresystems findet sich bei dem Grundsatz Individualisierbarkeit ( $\bar{x} = 2.41$ ); der höchste Wert bei einer anderen Software bei dem Grundsatz Steuerbarkeit ( $\bar{x} = 6.13$ ). Der ISONORM 9241/10-Gesamtmittelwert beträgt  $\bar{x} = 4.64$  ( $\bar{x}_{\text{min}} = 2.97$ ;  $\bar{x}_{\text{max}} = 5.75$ ;  $s = .65$ ;  $N = 41$  Programme).

Abbildung 3 zeigt Softwareprogramme geordnet nach ihrem ISONORM 9241/10-Gesamtwert. Eine nähere Betrachtung verdeutlicht, daß 34 der dort aufgeführten 41 Softwareprogramme einen ISONORM 9241/10-Gesamtwert von  $\bar{x} \geq 4.0$  erreichen und damit entsprechend der zugrundeliegenden Fragebogenskala im positiven Bereich angesiedelt sind. Allerdings erreichen lediglich 14 Programme einen ISONORM 9241/10-Gesamtwert von  $\bar{x} \geq 5.0$ . Etwa zwei Drittel (65.8 %) der untersuchten Softwareprogramme erreichen diesen Wert, der auf der Fragebogenskala einer Konnotation von „+“ (Plus Eins) entspricht, nicht.

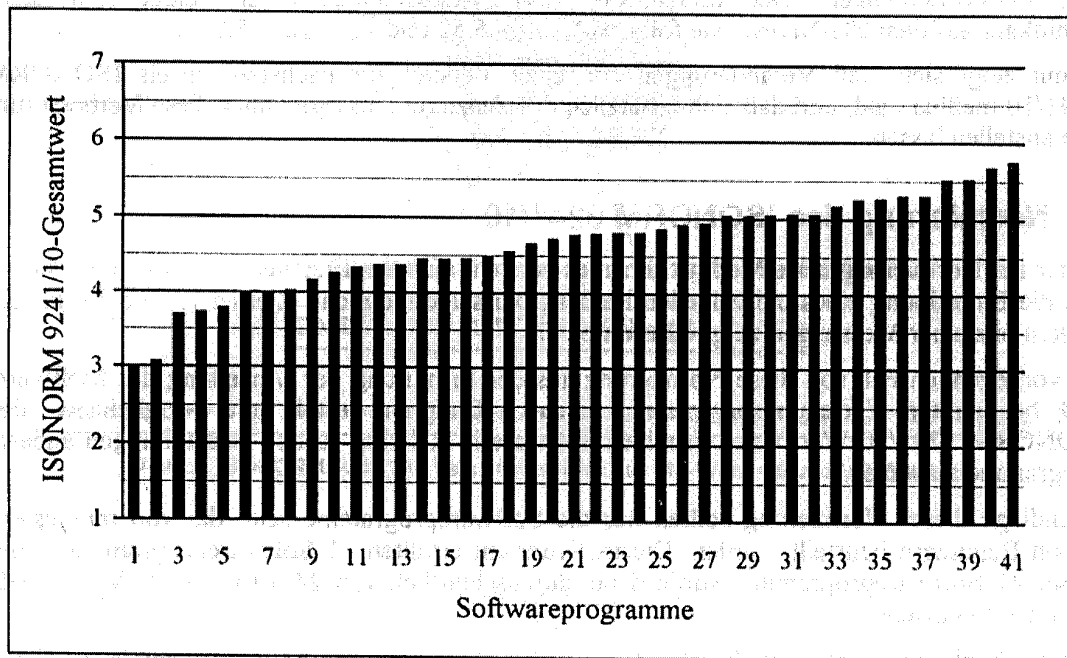


Abb. 3: Softwareprogramme geordnet nach ISONORM 9241/10-Gesamtwert

## 5 Zusammenfassung

Ziel dieses Beitrags war die Dokumentation der testtheoretische Güte des Benutzungsfragebogens ISONORM 9241/10. Zu diesem Zweck wurden verschiedene Untersuchungen zur Reliabilität, Validität und Normierung durchgeführt.

Zur Reliabilitätsbestimmung wurden die *Wiederholungsmethode* und die *Konsistenzanalyse* herangezogen. Aufgrund der signifikanten Stabilitäts-Koeffizienten konnte mittels der *Wiederholungsmethode* gezeigt werden, daß sich der ISONORM 9241/10 durch eine hohe Merkmalskonstanz auszeichnet. Anhand der *Konsistenzanalyse* zeigte sich die Leistungsfähigkeit des ISONORM 9241/10 als Meßinstrument, da die einzelnen Skalen des ISONORM 9241/10 gute innere Konsistenzen aufwiesen.

Im Rahmen der Validitätsbestimmung wurden Untersuchungen zur *kriterienbezogenen Validität* und zur *Konstruktvalidität* angestellt. Zur Überprüfung der *inneren kriterienbezogenen Validität* wurden die ISONORM 9241/10-Bewertungen mit den Bewertungen der beiden Benutzungsfragebögen QUIS [21] und BBD [22] korreliert. Es zeigte sich, daß der ISONORM 9241/10 zu vergleichbaren Ergebnissen gelangt, und gleichzeitig den Vorteil in sich birgt, eine ausdrückliche Operationalisierung von ISO 9241 Teil 10 darzustellen. Zur Überprüfung der



*äußeren kriterienbezogenen Validität* wurde mittels EVADIS II [9] überprüft, ob die Befragung durch Benutzer mit der Evaluation durch Experten korrespondiert. Es konnte gezeigt werden, daß die ISONORM 9241/10-Benutzungsanalysen einen signifikanten Bezug zu einer systematischen, leitfadenorientierten Expertenevaluation aufweisen. Zur Überprüfung der *Konstruktvalidität* wurden zwei Versionen eines Softwareprogramms mit und zwei Versionen eines Softwareprogramms ohne graphische Benutzungsschnittstelle herangezogen. Es zeigte sich, daß Veränderungen von einer Version zur nächsten mittels ISONORM 9241/10 meßbar sind, und daß sich zusätzlich differenzierte Aussagen über diese Verbesserungen anstellen lassen.

Die *Normierung* erlaubt für den ISONORM 9241/10 ein Bezugssystem herzustellen, das die Einordnung einer einzelnen Software in Relation zu den Beurteilungen anderer Programme ermöglicht.

Neben diesen Ergebnissen sollten auch Überlegungen zur *Ökonomie* und *Nützlichkeit* eines Meßverfahrens angestellt werden. Ein Meßverfahren ist dann *ökonomisch*, wenn es „eine kurze Durchführungszeit beansprucht, wenig Material verbraucht, einfach zu handhaben, als Gruppentest durchführbar, schnell und bequem auszuführen ist“ [7, S. 12]. Diese Kriterien werden durch den ISONORM 9241/10 ausnahmslos erfüllt. *Nützlich* ist ein Meßverfahren dann, wenn es ein Merkmal mißt, für dessen Untersuchung ein praktisches Bedürfnis besteht. Als direkte Operationalisierung der ISO 9241/10 erfüllt der ISONORM 9241/10 auch dieses Kriterium.

Die schriftliche Befragung von Benutzern kann und darf nicht die einzige Methode der Wahl sein, wenn eine Software umfangreich evaluiert werden soll. Aufgrund der Analyse einer Vielzahl von Evaluationsverfahren kann deswegen auch der Schluß gezogen werden, daß es die „beste“ Evaluationsmethode nicht gibt, sondern daß die jeweilige Eignung vom Gegenstand und Einsatzzweck der Evaluation bestimmt wird [20]. Jede Evaluationsmethode hat ihre spezifischen Vorteile und Schwerpunkte. Um ganzheitliche Aussagen bezüglich der software-ergonomischen Qualität von Benutzungsschnittstellen machen zu können, ist deshalb der Einsatz einer Kombination unterschiedlicher Evaluationsmethoden empfehlenswert [10]. Hier reicht das Spektrum von der heuristischen Evaluation [8], über die software-unterstützte, leitfadenorientierte Evaluation [11], bis hin zur handlungsorientierten Fehleranalyse [2,3,5,13,18], um nur einige Beispiele zu nennen.

Auf eine systematische und normbasierte Befragung durch Benutzer sollte jedoch auf keinen Fall verzichtet werden, da es z.B. vorkommen kann, daß Experten eine Software für effektiv und effizient halten, daß aber Benutzer mit der Software nicht zufrieden sind [1]. Mit dem ISONORM 9241/10 wurde ein reliables und valides Instrument entwickelt, das für diesen Zweck zum Einsatz kommen kann.

## Literaturverzeichnis

- [1] W. Dzida: Qualitätssicherung durch software-ergonomische Normen. In: E. Eberleh, H. Oberquelle & R. Oppermann (Hg.): Einführung in die Software-Ergonomie. DeGruyter, Berlin, 1994, 374-406.
- [2] M. Frese & D. Zapf (Hg.): Fehler bei der Arbeit mit dem Computer - Ergebnisse von Beobachtungen und Befragungen im Bürobereich. Huber, Bern, 1991.
- [3] M. Hassenzahl, J. Prümper & U. Sailer: Die Priorisierung von Problemhinweisen in der software-ergonomischen Qualitätssicherung. In diesem Band.
- [4] ISO 9241/10: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 10: Grundsätze der Dialoggestaltung. CEN, Brüssel, 1995.
- [5] A. Kensik, J. Prümper & M. Frese: Ergonomische Gestaltung von Software auf Grundlage handlungsorientierter Fehleranalysen. In: H.D. Böcker (Hg.): Software-Ergonomie '95 - Anwendungsbereiche lernen voneinander. Teubner, Stuttgart, 1995, 217-232.

- [6] A. Kinder: Testen und Bewerten von Software durch Benutzer. Justus-Liebig-Universität, Gießen, 1991.
- [7] G.A. Lienert & U. Ratz: Testaufbau und Testanalyse. Psychologie Verlags Union, München, 1994.
- [8] Nielsen, J. Usability Engineering. Cambridge: AP Professional, 1993.
- [9] R. Oppermann, B. Murchner, H. Reiterer & M. Koch: Software-ergonomische Evaluation. Der Leitfaden EVADIS II. De Gruyter, Berlin, 1992.
- [10] R. Oppermann & H. Reiterer. Software-ergonomische Evaluation. In: E. Eberleh, H. Oberquelle & R. Oppermann (Hg.): Einführung in die Software-Ergonomie. DeGruyter, Berlin, 1994, 335-371.
- [11] R. Oppermann, C. Wick, T. Geis, M. Koch, P. Lutz, J. Prümper, H. Reiterer & W.W. Strapetz: Der ISO 9241 Evaluator. In: Ergonomie & Informatik 27 (1996), 13-17.
- [12] J. Prümper: Benutzerorientierte, iterative Software-Entwicklung in der Praxis. In: W. Coy, P. Gorny, I. Kopp & C. Skarpelis (Hg.): Menschengerechte Software als Wettbewerbsfaktor. Teubner, Stuttgart, 1993, 630-647.
- [13] J. Prümper: Fehlerbeurteilungen in der Mensch-Computer Interaktion. Reliabilitätsanalysen und Training einer handlungstheoretischen Fehlertaxonomie. Waxmann, Münster, 1994.
- [14] J. Prümper: GUIs sind noch weit entfernt von einer optimalen Bewertung. ISO 9241 Teil 10 als Leitlinie für SW-Entwicklung und -Bewertung. In: Computerwoche 25 (1993), 29 & 46.
- [15] J. Prümper: Software-Evaluation based upon ISO 9241 Part 10. In: T. Grechenig & M. Tscheligi (Hg.): Human Computer Interaction. Springer, Berlin, 1993, 255-265.
- [16] J. Prümper: Viele Programme erfüllen die ISO-Norm 9241/10 ungenügend. Anwenderbeurteilung zeugt von Nachholbedarf. In: Computerwoche 21 (1994), 47-48.
- [17] J. Prümper & M. Anft: Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel. In: K.H. Rödiger (Hg.): Software-Ergonomie '93 - Von der Benutzungsoberfläche zur Arbeitsgestaltung. Teubner, Stuttgart, 1993, 145-156.
- [18] J. Prümper, T. Heinbokel & H.J. Küting: Virtuelle Prototypen als Werkzeuge zur benutzerzentrierten Produktentwicklung. Anwendung einer handlungstheoretischen Fehlertaxonomie auf reale und simulierte Oberflächen von Waschmaschinen. In: Zeitschrift für Arbeitswissenschaft 4 (1993), 160-167.
- [19] M. Rauterberg: Läßt sich die Gebrauchstauglichkeit interaktiver Software messen? Und wenn ja, wie? In: Ergonomie & Informatik 16 (1992), 3-18.
- [20] H. Reiterer: Ergonomische Kriterien für die menschengerechte Gestaltung von Bürosystemen, Anwendung und Bewertung. Universität Wien, Dissertation, 1990.
- [21] B. Shneiderman: Designing the user interface. Addison-Wesley, Reading MA, 1987.
- [22] P. Spinax: Arbeitspsychologische Aspekte der Benutzerfreundlichkeit von Bildschirmsystemen. ADAG, Zürich, 1987.

Quelle:

Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In R. Liskowsky, B.M. Velichkovsky & W. Wünschmann (Hrsg.), *Software-Ergonomie '97 – Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (S. 253-262). Stuttgart: Teubner.