



UNIVERSITÄT REGENSBURG

LEHRSTUHL FÜR INFORMATIONSWISSENSCHAFT

ANWENDERSTUDIE

EVALUIERUNG VOICE READER

1 Konzeption

1.1 Aufbau der Methodik

1.1.1 Variablendefinition

Unabhängige Variablen:

Testpersonen – Alter, Geschlecht, Berufstätigkeit, Erfahrung mit ähnlichen Programmen

Art der Informationsaufnahme - in drei Stufen eingeteilt – Lesen, Hören, Lesen & Hören

Abhängige Variablen (Messgrößen gem. ISO 9241-11 „Anforderungen an die Gebrauchstauglichkeit“):

Effektivität – Qualität der Aufgabenerledigung (Reproduktion eines Textes)

Effizienz – misst die Effektivität im Verhältnis zum benötigten Zeitaufwand.

Zufriedenheit mit der Anwendung – Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produkts

Zusätzlich wird auch die *Zeitersparnis* bei der Benutzung des Voice Reader gemessen.

1.1.2 Statistische Hypothesen

Alternativhypothesen (gerichtete Unterschiedshypothesen) H_1 :

1. *Die Verwendung des Voice Reader (bei Hören und Lesen & Hören) führt zur Steigerung der Qualität der Aufgabenerledigung (Effektivität).*
2. *Die Verwendung des Voice Reader (bei Hören und Lesen & Hören) führt zur Steigerung der Effizienz.*
3. *Durch die Verwendung des Voice Reader (bei Hören und Lesen & Hören) wird die Zufriedenheit der Benutzer gesteigert.*

Nullhypothesen (H_0):

1. *Die Verwendung des Voice Reader (bei Hören und Lesen & Hören) führt zu keiner Steigerung bzw. führt zu Verringerung der Qualität der Aufgabenerledigung (Effektivität).*
2. *Die Verwendung des Voice Reader (bei Hören und Lesen & Hören) führt zu keiner Steigerung bzw. führt zu Verringerung der Effizienz.*

3. *Durch die Verwendung des Voice Reader (bei Hören und Lesen & Hören) wird die Zufriedenheit der Benutzer nicht gesteigert bzw. wird verringert.*

1.1.3 Testanordnung

Mehr-Gruppen-Plan-Untersuchung, weil eine mehrfach gestufte unabhängige Variable und mehrere abhängige Variablen vorhanden sind.

- Optimaler Stichprobenumfang – Stichprobenumfänge sind optimal, wenn sie bei gegebenen Irrtumswahrscheinlichkeit, Teststärke und Effektgröße eine eindeutige Entscheidung über die Gültigkeit von H_0 oder H_1 sicherstellen.

Bei einer Irrtumswahrscheinlichkeit $\alpha = 0,05$, Teststärke $1 - \beta = 0,80$ und einer Effektgröße $\varepsilon = 0,35$ wurde ein optimaler Stichprobenumfang von $n = 27$ für jede Stufe der unabhängigen Variable bestimmt. Bei dieser Untersuchung hat jede Versuchsperson alle drei Arten der Informationsaufnahme ausprobiert, deshalb nahmen insgesamt 27 Probanden daran teil.

Zum Vermeiden von Lerneffekten und anderen Einflüssen wurden die Texte und die Art der Informationsaufnahme in folgender Reihenfolge rotiert:

Reihenfolge:	1	2	3
Proband 1	Lesen Text 1	Hören Text 2	Lesen und Hören Text 3
Proband 2	Hören Text 3	Lesen und Hören Text 1	Lesen Text 2
Proband 3	Lesen und Hören Text 2	Lesen Text 3	Hören Text 1
Proband 4	Lesen Text 2	Hören Text 3	Lesen und Hören Text 1
Proband 5	Hören Text 1	Lesen und Hören Text 2	Lesen Text 3
Proband 6	Lesen und Hören Text 3	Lesen Text 1	Hören Text 2
Proband 7	Lesen Text 3	Hören Text 1	Lesen und Hören Text 2
Proband 8	Hören Text 2	Lesen und Hören Text 3	Lesen Text 1
Proband 9	Lesen und Hören Text 1	Lesen Text 2	Hören Text 3

Dieselbe Reihenfolge wurde noch zweimal bis 27 wiederholt.

1.1.4 Berechnung der Messgrößen

- Effektivität: Die Effektivität wurde anhand des Maßes *Recall* bestimmt. Der *Recall* zeigt, wie vollständig eine Frage beantwortet wurde, und reicht von 0 bis 1 (0 - 100%).

Ein Beispiel zum Text 3:

Frage 6: „Wie heißt die Frau des dritten Mannes und wo arbeitet sie?“

Antwort: „Erna, sie arbeitet in der Praxis“

Wenn die Antwort vollständig ist, wird sie mit 1 benotet (100%), und wenn nur eine der beiden Aussagen angegeben ist, wird die Antwort mit 0,5 bewertet (50%). Danach wird der Mittelwert von allen 10 Fragen zum jeweiligen Text berechnet.

- Effizienz: die Effizienz wurde anhand folgender Formel bestimmt:

$$\text{Effizienz} = \frac{\text{Effektivität}}{\text{Bearbeitungsdauer}}$$

- Zufriedenheit mit der Anwendung: Die Zufriedenheit der Probanden mit Voice Reader wurde anhand der subjektiven Einschätzung für den Einsatz des Programms und der Verbesserungsvorschläge gemessen.
- Zeitersparnis:

$$\text{Zeitersparnis} = \frac{\text{Bearbeitungsdauer(Lesen\&Hören)} - \text{Bearbeitungsdauer(Lesen)}}{\text{Bearbeitungsdauer(Lesen)}}$$

$$\text{Zeitersparnis} = \frac{\text{Bearbeitungsdauer(Lesen\&Hören)} - \text{Bearbeitungsdauer(Hören)}}{\text{Bearbeitungsdauer(Hören)}}$$

$$\text{Zeitersparnis} = \frac{\text{Bearbeitungsdauer(Hören)} - \text{Bearbeitungsdauer(Lesen)}}{\text{Bearbeitungsdauer(Lesen)}}$$

1.1.5 Pretest

Es wurde zunächst ein Pretest durchgeführt. Dabei stellte sich heraus, dass vier Minuten zum einmaligen Lesen des Textes völlig ausreichen. Der Fragebogen wurde danach nochmals überarbeitet.

1.2 Testablauf

1.2.1 Begrüßung, Programmvorstellung und Anweisungen

1.2.2 Aufgabenbearbeitung

- Jede Versuchsperson bekommt drei Texte – einen zum Lesen, einen zum Hören und einen zum Lesen & Hören gleichzeitig (Rotation der Texte und der Art der Informationsaufnahme zum Vermeiden von Lerneffekten)
- Jeder Text darf nur einmal gelesen bzw. gehört bzw. gelesen & gehört werden.
- Zum Lesen haben die Probanden ausreichend Zeit (jedoch max. 4 Minuten) zur Verfügung. So wird sichergestellt, dass die Texte nicht mehr als einmal gelesen werden können.
- Bei der Hörpräsentation wird der Bildschirm ausgeschaltet, damit die Probanden nicht abgelenkt werden.
- Beantworten von 10 Fragen zum jeweiligen Text
- Es wird notiert, wie viel Zeit jeder Proband fürs Beantworten der Fragen zu jedem Text benötigt.
- Die Fragen zu den Texten sind schriftlich zu beantworten.

1.2.3 Ausfüllen des Fragebogens

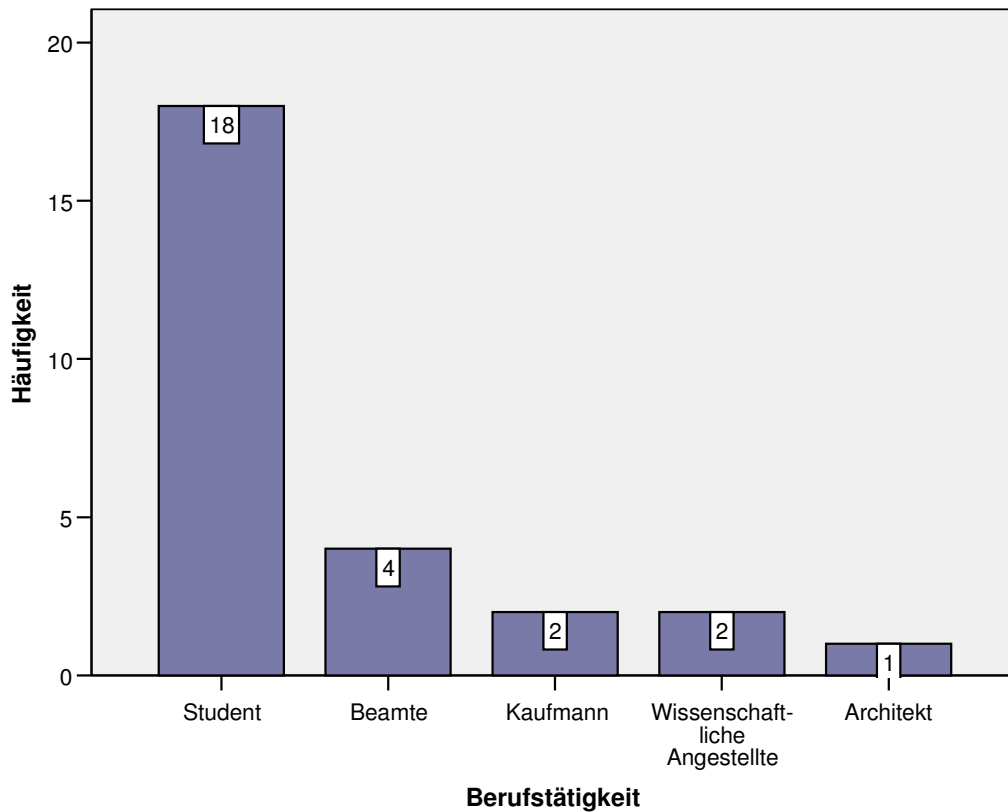
Fragen zu Benutzerdaten, persönlichen Präferenzen und Eindrücken bei der Benutzung des Programms, sowie weitere Fragen zum Erfassen von möglichen Relationen zwischen abhängigen und unabhängigen Variablen.

Der Fragebogen ist schriftlich auszufüllen.

2 Der Test

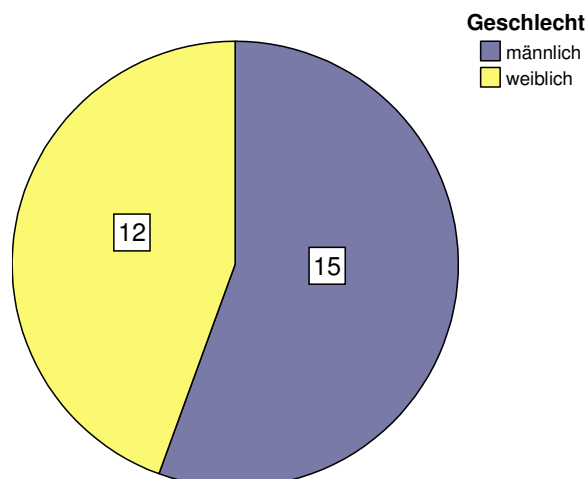
Der Test wurde im Laborraum der Informationswissenschaft an der Universität Regensburg durchgeführt. Der Test dauerte ca. 30 Minuten pro Proband. Es nahmen 27 Personen teil. Das folgende Diagramm zeigt die Berufstätigkeit der Probanden:

Abb.1: Berufstätigkeit der Probanden



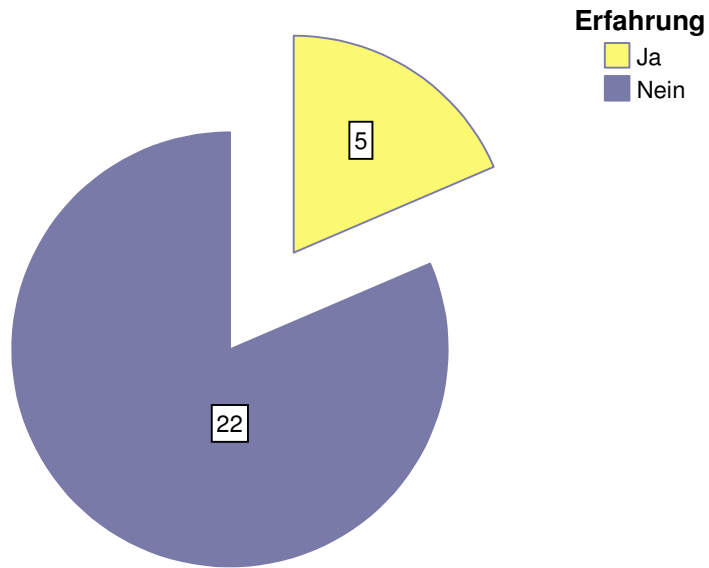
Die weiblichen und männlichen Probanden waren relativ gleich verteilt - von den 27 Personen waren 12 weiblich und 15 männlich:

Abb.2: Geschlecht



Erfahrung mit anderen Vorlese-Programmen hatten 5 Personen, 22 haben zum ersten Mal mit einem solchen Tool gearbeitet.

Abb.3: Erfahrung

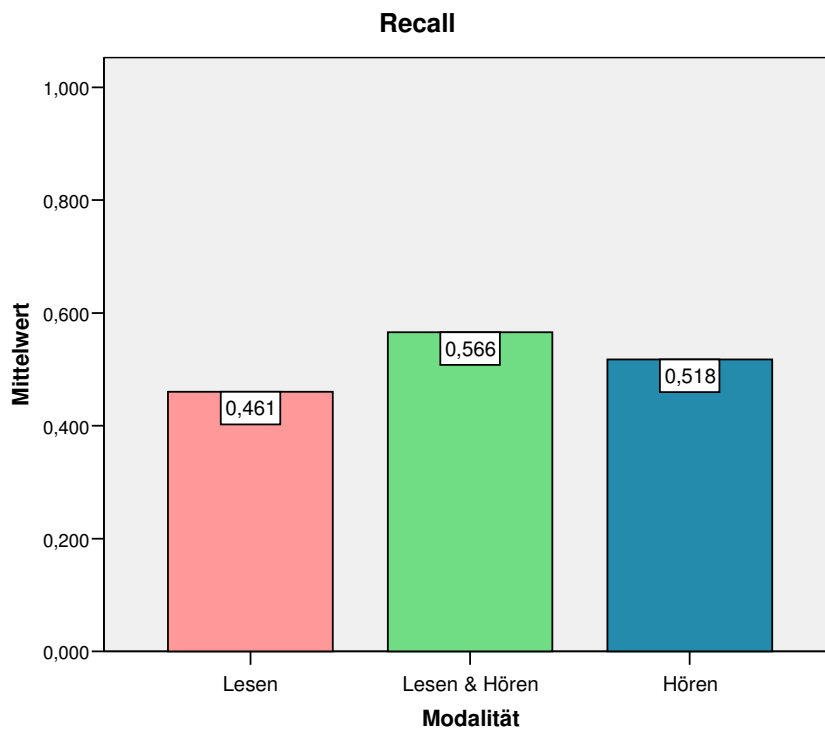


3 Auswertung, Analyse und Präsentation der Daten

3.1 Effektivität

Im folgenden Diagramm (Abb. 4) sind die Mittelwerte des Recalls für die drei Arten der Informationsaufnahme dargestellt. Am besten hat „Lesen & Hören“ mit 0,566 abgeschnitten, gefolgt von „Hören“ mit 0,518 und „Lesen“ mit 0,461.

Abb.4: Effektivität



Um präzisere Erkenntnisse über die Richtigkeit der Nullhypothese zu gewinnen, wurde eine einfaktorische Varianzanalyse (ANOVA) durchgeführt. Der Varianzanalyse liegt die Überlegung zugrunde, dass sich die gesamte Streuung der betrachteten Variablen in der vorliegenden Stichprobe gedanklich aufteilen lässt in die Streuung innerhalb der einzelnen Fallgruppen einerseits und die Streuung zwischen den Gruppen andererseits. Der F-Wert wird durch den Vergleich der Quadratsumme innerhalb der Gruppen mit der Quadratsumme zwischen den Gruppen berechnet. Da der empirisch ermittelte F-Wert (3,158) größer als der kritische (3,113) ist, wird die Nullhypothese auf dem 5%-Niveau verworfen. Es wird dann die Alternativhypothese angenommen, dass die Verwendung des Voice Reader (bei Hören und Lesen & Hören) zur Steigerung der Effektivität führt.

Abb.5: ANOVA Effektivität

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,151	2	,075	3,158	,048
Within Groups	1,859	78	,024		
Total	2,010	80			

Der Duncan-Test liefert auf dem 5%-Signifikanzniveau zwei homogene Untergruppen. Signifikant unterschiedlich sind die Ergebnisse von „Lesen“ auf der einen Seite und „Lesen & Hören“ auf der anderen Seite, die daher zwei getrennten „Untergruppen“ zugeordnet wurden. Der Mittelwert von „Hören“ wird beiden Gruppen zugewiesen, da der Unterschied weder zum Mittelwert von „Lesen“ und noch zum Mittelwert von „Lesen & Hören“ signifikant ist.

Abb.6: Duncan-Test Effektivität

Effektivität

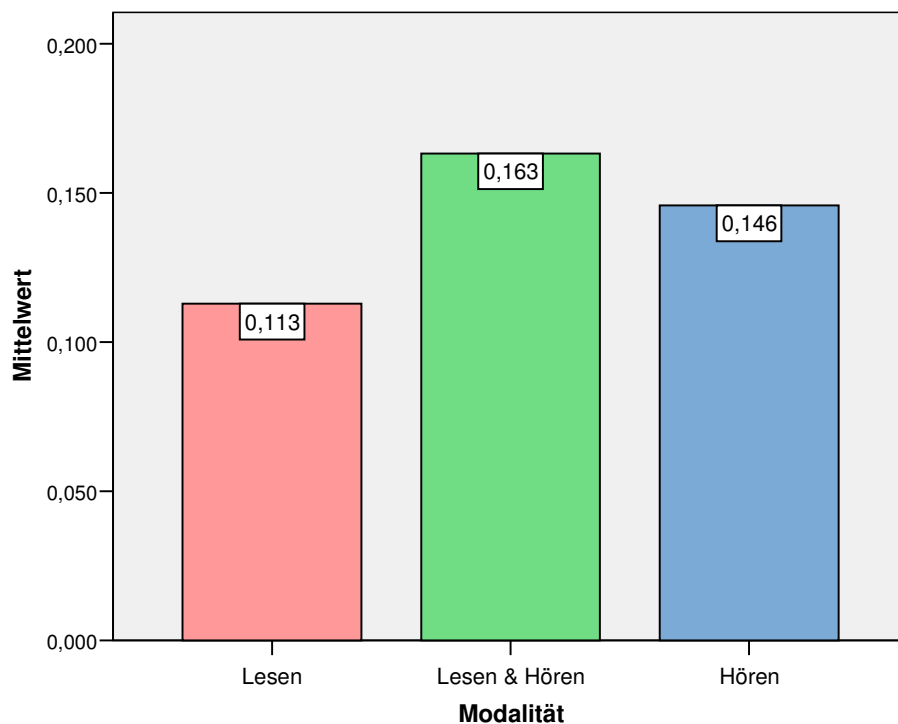
Modalität	N	Subset for alpha = .05	
		1	2
Lesen	27	,46056	
Hören	27	,51756	,51756
Lesen & Hören	27		,56604
Sig.		,179	,252

Means for groups in homogeneous subsets are displayed.
A Uses Harmonic Mean Sample Size = 27,000.

3.2 Effizienz

Die höchste Effizienz¹ hat das „Lesen & Hören“ (0,163) erzielt, gefolgt von „Hören“ (0,146) und schließlich „Lesen“ (0,113).

Abb.7: Effizienz



¹ Effizienz = Recall / Bearbeitungsdauer

Da der empirisch ermittelte F-Wert (4,496) auch hier größer als der kritische (3,113) ist, wird die Alternativhypothese angenommen, dass die Verwendung des Voice Reader (bei Hören und Lesen & Hören) zur Steigerung der Effizienz führt.

Abb.8: ANOVA Effizienz
Effizienz

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,035	2	,018	4,496	,014
Within Groups	,306	78	,004		
Total	,342	80			

Nach dem Duncan-Test unterscheidet sich der Effizienzmittelwert von „Lesen & Hören“ signifikant vom Effizienzmittelwert von „Lesen“. Der Unterschied zum Effizienzmittelwert von „Hören“ ist dagegen nicht signifikant. Auch hier wird der Mittelwert von „Hören“ beiden Gruppen zugewiesen, da der Unterschied weder zum Mittelwert von „Lesen“ und noch zum Mittelwert von „Lesen & Hören“ signifikant ist.

Abb.9: Duncan-Test Effizienz

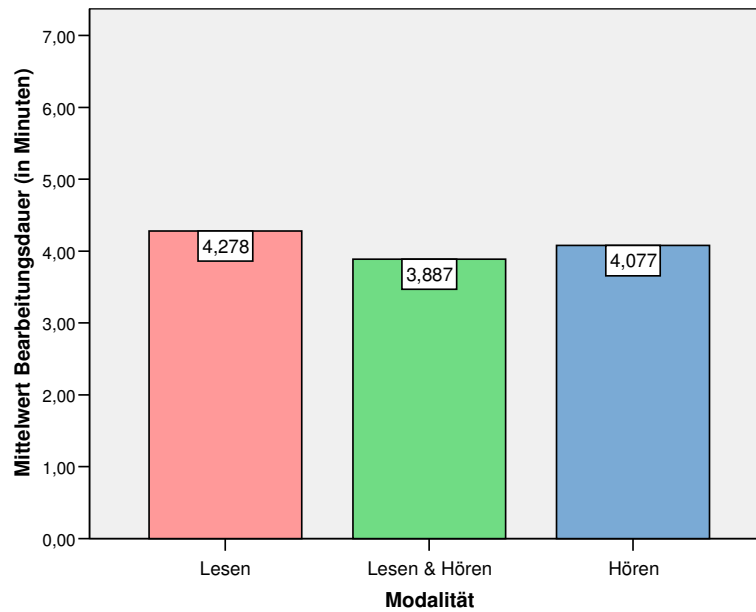
Effizienz

Modalität	N	Subset for alpha = .05	
		1	2
Lesen	27	,11289	
Hören	27	,14581	,14581
Lesen & Hören	27		,16326
Sig.		,057	,310

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 27,000.

Allein die Bearbeitungsdauer zeigt, dass die Probanden am wenigsten Zeit für die Beantwortung der Fragen bei „Lesen & Hören“ gebraucht haben (3,89 Minuten im Mittel), gefolgt von „Hören“ (4,08 Minuten) und „Lesen“ (4,28 Minuten).

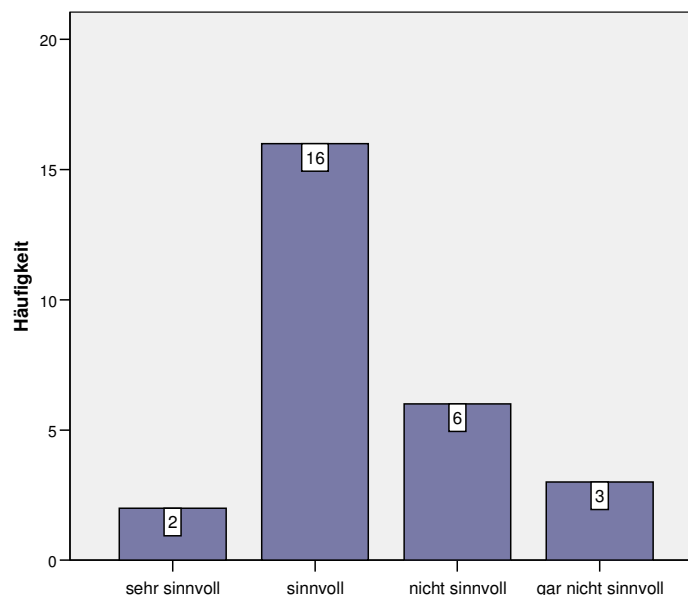
Abb.10: Bearbeitungsdauer



3.3 Zufriedenheit mit der Anwendung

Die Zufriedenheit der Probanden mit Voice Reader wurde anhand der subjektiven Einschätzung für den Einsatz des Programms und der Verbesserungsvorschläge gemessen. Auf die Frage, für wie sinnvoll die Probanden für sich selbst den Einsatz solcher Vorlese-Programme halten, haben 2 mit „sehr sinnvoll“ und 16 Personen mit „sinnvoll“ geantwortet. 6 dagegen halten es für „nicht sinnvoll“ und 3 Personen für „gar nicht sinnvoll“.

Abb.11: Für wie sinnvoll halten Sie für sich selbst den Einsatz des Voice Reader?



66% der Probanden (18) haben das Programm als „sehr sinnvoll“ und „sinnvoll“ bewertet. Daher wird auch hier die Alternativhypothese angenommen, die besagt, dass durch die Verwendung des Voice Reader die Zufriedenheit der Benutzer gesteigert wird.

4 Zusammenfassung

Abschließend lässt sich festhalten, dass der Einsatz des Vorlese-Programms Voice Reader die Bearbeitungszeit von Dokumenten reduzierte und das Textverständnis steigerte. Die Effizienz lag beim Hören um 29% höher als beim Lesen. Das beste Ergebnis erzielten Probanden, die einen Text gleichzeitig gelesen und angehört haben. Hier ergab sich im Vergleich zum Lesen eine Verbesserung um 44%.