

# TestASS: Eine Methode zur Testaufwandsschätzung

*Ulrike Dowie, Georg Herzwurm*

Universität Stuttgart, Lehrstuhl für allg. BWL und Wirtschaftsinformatik II

ulrike.dowie@gmx.de, herzwurm@wi.uni-stuttgart.de

## **Zusammenfassung:**

*Im Rahmen einer Forschungsarbeit zur Ermittlung der Einflussfaktoren des Testaufwands wurde eine Methode entwickelt und erprobt, mit Hilfe derer aus der Menge potenzieller Einflussfaktoren die in einer bestimmten Organisation relevanten Einflussfaktoren identifiziert werden können. Zusammen mit der Auswahl und Integration geeigneter Schätzmethode kann so ein organisationsspezifisches Vorgehen zur Schätzung des Testaufwands hergeleitet werden. Dieser Beitrag stellt die Methode TestASS (Test-Aufwands-Schätzung für Standardsoftware) vor.*

## **Schlüsselbegriffe**

*Testaufwand, Aufwandsschätzung, organisationsspezifisches Modell*

## **1 Einleitung**

Wie viel Testaufwand in einem Softwareentwicklungsprojekt zu betreiben ist, hängt neben dem festgelegten Ziel der Tests (z. B. eine bestimmte, erwartete Anzahl an Fehlern zu finden, vgl. [14], S. 95) von sehr vielen Einflussfaktoren ab ([10], S. 56).

Die Menge der potenziellen Einflussfaktoren lässt sich jedoch reduzieren, wenn der Testaufwand in nur einer Organisation geschätzt werden soll ([12], S. 85). Insbesondere der Einfluss, den das Vorgehensmodell auf den Testaufwand ausübt ([6], S. 59), kann außer Betracht bleiben, wodurch sich die Modellierung und Schätzung des Testaufwands deutlich vereinfachen.

Im Folgenden wird eine Methode beschrieben, um aus den möglichen Einflussfaktoren des Testaufwands die in einer Organisation Relevanten auszuwählen und ein organisationsspezifisches, anwendbares Modell zur Testaufwandsschätzung zu entwickeln.

## **2 Methodischer Ansatz**

Die Entwicklung einer Methode entspricht einer häufig angewendeten praktischen Forschungsform, wenn zur Erreichung des Forschungsziels eine Konstruktionsstrategie verfolgt wird. Die folgenden drei Elemente stellen die Säulen der Methode dar:

**Datensammlung:** Um die lokal, d. h. in einer Organisation wichtigen Einflussfaktoren des Testaufwands zu ermitteln, müssen die Zusammenhänge zwischen Testaufwand und möglichen Einflussfaktoren anhand mehrerer Projekte beobachtet und analysiert werden. Die hierfür nötige Datenbasis bieten historische Projektdaten, d. h. Daten zu bereits abgeschlossenen Projekten. Als Informationsquellen dienen einerseits Datenbanken, andererseits Projektmitarbeiter (insbesondere Projektleiter und Qualitätssicherungsbeauftragter) und Experten der Qualitätssicherung in der jeweiligen Organisation.

**Integration qualitativer und quantitativer Daten:** Sowohl quantitative (z. B. Projektdauer) als auch qualitative (z. B. Qualität der Spezifikation) Daten müssen zum Einsatz kommen ([1], S. 760). Während Basili und Rombach dies mit fehlendem Verständnis von Zusammenhängen und mangelnder Toolunterstützung begründen (vgl. ebenda), sind außerdem sowohl in empirischer Forschung als auch in der Praxis Kosten und Nutzen einer Messung abzuwägen. Zusätzlicher Aufwand, um quantitative Daten zu erheben, kann nur durch höhere Zuverlässigkeit und Genauigkeit gerechtfertigt werden, die jedoch nicht in jeder Organisation und zu jedem Merkmal gegeben sind (vgl. hierzu Abschnitt 4, Ergebnisse).

**Datenauswertung:** Es werden grafische, statistische und interpretative Analysen durchgeführt, die sich teils auf einzelne Projekte beziehen (z. B. interpretative Analyse besonders erfolgreicher Projekte), teils auf alle Projekte (z. B. Korrelations- und multiple Regressionsanalysen).

### 3 Bezug zum Stand der Forschung

Für die Schätzung des Entwicklungsaufwands in der Softwareentwicklung gibt es eine Reihe von Modellen. Das wohl bekannteste Beispiel ist CoCoMo, das „Constructive Cost Model“ von Boehm u. a., das im Jahr 2000 überarbeitet als CoCoMo II veröffentlicht wurde [2]. Aber auch zur Schätzung des Testaufwands finden sich in der Fachliteratur mehrere Modelle bzw. Methoden (z. B. [3], [8] sowie [9]). Diese Ansätze sind hinsichtlich des Gültigkeitsbereichs nicht eingeschränkt, basieren allerdings teilweise auf Annahmen, die nicht in jeder Organisation gegeben sind. Beispiele hierfür sind die Annahme, die Anzahl der im Softwareprodukt enthaltenen Fehler sei zum Testbeginn bekannt ([3], S. 784), und dass Testfälle hinsichtlich des Durchführungsaufwands vergleichbar sind ([9], S. 31). Außerdem fehlen häufig genaue Beschreibungen, wie vorzugehen ist, um den jeweiligen Ansatz zur Aufwandsschätzung anzuwenden. Beispielsweise berücksichtigt Nageswaran die Entwicklungs- und Testumgebung als Einflussfaktor und einen Zuschlag für die Projektkomplexität, ohne anzugeben, wie sie jeweils zu messen sind ([8], S. 4f). Im Modell von Cangussu u. a. wird die Qualität der Testphase als Einflussfaktor betrachtet, jedoch nicht beschrieben, wie der Faktor quantifiziert werden kann ([3], S. 790).

Anstatt eines uneingeschränkt anwendbaren Modells des Testaufwands wird daher in dieser Arbeit eine Methode präsentiert, die dank weniger Prämissen breit einsetzbar ist und mittels derer organisationsspezifische Modelle des Testaufwands entwickelt werden können. Die Schritte dieser Methode werden im folgenden Abschnitt im Detail beschrieben.

## **4 Herleitung eines lokalen Testaufwandsmodells: Die Methode TestASS**

### **4.1 Gültigkeitsbereich**

Aufgrund des hohen Verbreitungsgrads des V-Modells ist die Methode für Organisationen und Entwicklungsprojekte zugeschnitten, die dieses Vorgehensmodell einsetzen.

Die Methode dient der Aufwandsschätzung in Projekten zur Standardsoftwareentwicklung. Im Gegensatz zur Auftragsentwicklung sind nicht kundenspezifische Anforderungen und vertragliche Vereinbarungen mit einem Kunden maßgeblich, sondern die Anforderungen eines organisationsinternen Auftraggebers. Diese Einschränkung schließt Migrations- und Einführungsprojekte ebenfalls vom Gültigkeitsbereich der Methode aus.

### **4.2 Schritte der Methode**

Die entwickelte Methode besteht aus folgenden fünf Schritten: Prüfung der Voraussetzungen zur Anwendung, Sammlung historischer Projektdaten, Bestimmung der lokal wichtigen Einflussfaktoren, Auswahl und Integration geeigneter Schätzmethoden und Aktualisierung des Vorgehens.

#### **Schritt 1: Voraussetzungen für die Anwendbarkeit prüfen**

Die wichtigste Voraussetzung ist, dass der Testaufwand abhängig vom Entwicklungsaufwand geplant wird. Der Grund hierfür ist, dass mit dem Umfang eines entwickelten oder geänderten Softwareprodukts auch der Umfang zu prüfender Artefakte (wie Spezifikation, Design und Code) zunimmt. Daher wird mittels des lokalen Modells Testaufwand nicht als absoluter Wert, sondern relativ zum Entwicklungsaufwand geschätzt.

Der Einfluss der verwendeten Testmethoden auf den Testaufwand wird nicht modelliert, weshalb vorausgesetzt wird, dass sich die Testmethoden zwischen Projekten hinreichend ähnlich sind. So lässt sich die Wirkung anderer messbarer Merkmale klarer erkennen.

Des Weiteren wird vorausgesetzt, dass Informationen zu abgeschlossenen Projekten verfügbar sind. Nicht nur Datenbanken und Projektdokumentation müssen zugänglich sein, sondern auch Ansprechpartner zu jedem der abgeschlossenen Projekte.

## **Schritt 2: Sammlung historischer Projektdaten**

In der Fachliteratur ist die Notwendigkeit einer organisationsspezifischen Datensammlung zu abgeschlossenen Projekten für die Aufwandsschätzung unstrittig (z. B. [13], S. 174, [4], [5], S. 3). Solche historischen Projektdaten lassen Zusammenhänge zwischen Projektmerkmalen und der Anzahl im Produktivbetrieb gefundener Fehler erkennen ([15], S. 74), dienen zur Evaluierung und zum Vergleich von Projekten und zur Modellierung weiterer organisationsspezifischer Zusammenhänge.

Um die Ausprägung der potenziellen Einflussfaktoren in jedem Projekt zu erfassen und Zusammenhänge zwischen den Faktoren und dem Testaufwand zu analysieren, müssen messbare Kennzahlen für alle zu untersuchenden Merkmale abgeleitet werden. Folgende messtheoretische Prinzipien kommen hierbei zur Anwendung:

1. Organisationsspezifische Messgrößen und Messinstrumente
2. Eindeutige und präzise Definitionen der Messgrößen
3. Subjektive und objektive Messgrößen und Messinstrumente
4. Mehrdimensionale Messung

Als Informationsquellen dienen in erster Linie Interviews mit dem Projektleiter, dem Qualitätssicherungsverantwortlichen, dem Produktarchitekten und einem Vertreter des Auftraggebers zu jedem Projekt. Darüber hinaus werden Datenbanken herangezogen, in denen beispielsweise der Projektaufwand und gefundene Fehler erfasst wurden, sowie Projektdokumentation wie z. B. Reviewprotokolle und Präsentationen anlässlich von Meilensteinen im Projekt.

Die Projekte werden in Form einzelner Fallstudien untersucht und sind nicht in ihrer Gesamtheit als Stichprobe zu verstehen.

## **Schritt 3: Bestimmung lokal wichtiger Einflussfaktoren des Testaufwands**

Die Bestimmung der lokal wichtigsten Einflussfaktoren erfolgt in sechs Schritten:

1. Ermittlung projektspezifischer Testaufwandstreiber

Hierzu wird der jeweilige Qualitätssicherungsverantwortliche in den zu analysierenden Projekten nach den projektspezifischen Testaufwandstreibern gefragt. Anhand einer Vorschlagsliste sind potenzielle Treiber des Testaufwands (z. B. Genauigkeit und Dynamik der Anforderungen, Engagement der Tester, Aufwandsverteilung auf Teststufen) hinsichtlich ihres Einflusses im Projekt zu bewerten und ggf. um weitere Aufwandstreiber zu ergänzen.

2. Identifikation von Vergleichskriterien hinsichtlich des Testaufwands

Mitarbeiter der Qualitätssicherungsabteilung, die über mehrjährige Erfahrung mit ihren Aufgaben in dieser Organisation verfügen, und Projektleiter werden befragt,

welche Projektmerkmale relevant seien zum Vergleich des Testaufwands zwischen Projekten. Die resultierende Rangliste gibt neben Kriterien, anhand derer die Projekte gruppiert werden können, weiteren Aufschluss über relevante Einflussfaktoren des Testaufwands.

### 3. Korrelationsanalysen

Für je einen potenziellen Einflussfaktor und den relativen Testaufwand, d. h. standardisiert in Bezug zum Entwicklungsaufwand, wird die statistische Korrelation bestimmt. Zusätzlich werden diese Zusammenhänge anhand von Streudiagrammen analysiert, um irreführende Einflüsse beispielsweise von Ausreißern zu erkennen. Besonders wichtig ist es, festgestellte Korrelationen anschließend anhand der Projektinformationen zu interpretieren und alternative Erklärungen für die Analyseergebnisse zu prüfen, um Fehlschlüsse hinsichtlich der Einflüsse auf den Testaufwand zu vermeiden.

### 4. Einzelfallanalysen

Das Ergebnis jedes Projekts hinsichtlich des erreichten Qualitätsziels und die Höhe des angefallenen Testaufwands werden anhand der erfassten Projektdaten interpretiert. Beide Ergebnisse werden bei der Analyse gemeinsam betrachtet, weil der mittels Modell zu schätzende Testaufwand ökonomischem Kalkül entsprechen soll, d. h. nur insoweit gerechtfertigt ist, als sich dadurch das angestrebte Qualitätsziel besser erreichen lässt. Andererseits ist der betriebene Testaufwand nur eine von mehreren Determinanten der Qualitätszielerreichung, weshalb auch andere Erklärungen für das jeweils erreichte Ziel zu betrachten sind.

### 5. Multiple Regressionsanalyse

Dieser Analyseschritt dient einerseits dazu, die bisherigen Erkenntnisse hinsichtlich der Einflussfaktoren zu bestätigen oder zu widerlegen. Andererseits liefert die Regressionsanalyse als Ergebnis ein mathematisch spezifiziertes Modell zur Testaufwandsschätzung.

Die Faktoren, die sich in den vorangegangenen Schritten als Einflussfaktoren des Testaufwands darstellten, werden zur Ermittlung einer Regressionsgleichung herangezogen, deren abhängige Variable der relative Testaufwand ist. Bei der Analyse werden verschiedene Kombinationen von unabhängigen Variablen berücksichtigt, die jeweils keine Multikollinearität der beteiligten Variablen aufweisen. Wie viele Variablen gemeinsam modelliert werden, hängt von der Anzahl der Projektdatensätze ab (als Faustregel werden 10 bzw. 15 Datensätze je modellierte unabhängige Variable genannt, vgl. [7], S. 1006 bzw. [11], S. 81).

Als Gütemaße der Regressionsgleichung werden das korrigierte Bestimmtheitsmaß, das Signifikanzniveau und der Standardfehler der Schätzung betrachtet. Unter mehreren Gleichungen wird diejenige ausgewählt, die ein relativ hohes Bestimmtheitsmaß mit einem niedrigen Signifikanzniveau und einem geringen Standardfehler verbindet. Lässt sich anhand der drei Gütemaße keine Gleichung ein-

deutig als beste bestimmen, so ist abzuwägen, wie hoch der Standardfehler sein darf, damit potenzielle Modellnutzer die Modellergebnisse akzeptieren.

Schließlich sind die ermittelten Koeffizienten der Variablen zu prüfen, insbesondere hinsichtlich des Vorzeichens und des Signifikanzniveaus, und die Voraussetzungen linearer Regression, vor allem, ob die unabhängigen Variablen der ausgewählten Gleichung linear mit dem Testaufwand verknüpft sind.

Das Ergebnis dieses Schritts ist ein mathematisch spezifiziertes, lokales Modell zur Testaufwandsschätzung. Extrahiert man daraus nur die Variablen (ohne Koeffizienten), was sich insbesondere anbietet, wenn deren Einfluss zwar plausibel und maßgeblich erscheint, die Prüfungen der Gleichung jedoch eine geringe Modellqualität vermuten lassen, erhält man ein lokales Modell in nicht-parametrisierter Form.

## 6. Validierung

Die Validierung des Modellierungsergebnisses erfolgt auf drei Arten: anhand der Beurteilung durch Experten innerhalb der Organisation, anhand von weiteren Projektdaten und in Pilotprojekten.

Zunächst werden die Ergebnisse der Regressionsanalyse, d. h. das lokale Modell zur Testaufwandsschätzung in spezifizierter und nicht-parametrisierter Form, einer Runde von Qualitätssicherungsexperten und potenziellen Nutzern des Modells vorgestellt und diskutiert. Ziel des Austauschs ist es, zu einem Konsens hinsichtlich der Beurteilung des Modells zu gelangen. Darüber hinaus werden Daten zu weiteren abgeschlossenen Projekten gesammelt, deren Testaufwand mittels Modell geschätzt wird. Die Abweichung des tatsächlichen vom geschätzten Testaufwand wird ermittelt und interpretiert, ob das Modell dahingehend verbessert werden kann, ein genaueres Schätzergebnis zu liefern. Schließlich zeigt die Anwendung des Modells in Pilotprojekten, welche Form (mathematisch spezifiziert oder nicht-parametrisiert) bevorzugt wird und welche Vor- und Nachteile die Nutzer bei der Testaufwandsschätzung anhand des Modells sehen.

### Schritt 4: Auswahl und Integration der Schätzmethode

Um die Stärken verschiedener Schätzmethode zu verbinden und gleichzeitig ihre jeweiligen Schwächen abzumildern, sind mehrere Schätzmethode auszuwählen. Das Vorgehen wird in folgender Tabelle 1 erläutert.

Nr.	Schritt	Erläuterung bzw. Beispiel	Ergebnis
1	Gültigkeitsbereich und Stakeholder der Testaufwands-	Erfolgt die Auswahl für ein Projekt, Programm, Organisation? Mögliche Stakeholder: Projektleiter, Qualitätssicherungs- und Budgetver-	Beteiligte der Schätzmethodeauswahl

Nr.	Schritt	Erläuterung bzw. Beispiel	Ergebnis
	schätzung (TAS) bestimmen	antwortliche,...	
2	Ziele und Rahmenbedingungen der TAS		
2.1	Ziele der TAS festlegen und gewichten	Mögliche Ziele: Ressourcen-, Zeitplanung, Projektdurchführung?; dabei verschiedene Nutzer (Testkoordinator, Projektleiter, Kunde,...) berücksichtigen	Gewichtete Ziele der TAS
2.2	Rahmenbedingungen der TAS identifizieren und gewichten	Beispiele: verfügbare Zeit, personelle und finanzielle Ressourcen für die Schätzung	Gewichtete Rahmenbedingungen der TAS
3	Verfügbare Schätzmethoden analysieren	Voraussetzungen, Stärken und Schwächen je Methode zusammenstellen	Methodenbeschreibung zur Beurteilung der Zielerfüllung
4	Zielerfüllung der Schätzmethoden bestimmen	Wie gut ist jede Methode geeignet, die einzelnen Ziele und Rahmenbedingungen zu erfüllen?	Matrix, die die Erfüllungsgrade jeder Schätzmethode hinsichtlich aller Ziele und Rahmenbedingungen enthält
5	Mehrere geeignete Schätzmethoden auswählen	Erfüllungsgrad multipliziert mit Gewicht des Ziels/der Rahmenbedingungen für alle Ziele und Rahmenbedingungen je Methode summieren; die zwei oder drei Methoden mit höchster Punktzahl auswählen	Rangfolge der Methoden
6	Ausgewählte Schätzmethoden integrieren	Mechanisch/vorab festgelegt oder freiwillig/interaktiv	Vorgehen zur Anwendung der ausgewählten Schätzmethoden

**Tabelle 1:** Auswahl geeigneter Methoden zur Testaufwandsschätzung (TAS)

Ein Beispiel für eine Matrix, die im Rahmen der Auswahl erstellt und genutzt wird, zeigt folgende Abbildung.

Ziele u. Rahmenbedingungen d. Schätzung	Schätzmethode	Schätzmethoden												
		parametr. Modell (Regr. analyse)	parametr. Modell (optimal subset sel.)	CoCoMo	Classification and Regression Trees	Analogieschluss informell	Indiv. Expertenschätzung: ad hoc	Indiv. Expertenschätzung: formalisiert	Gruppe von Experten: ad hoc	Gruppe von Experten: formalisiert	künstl. Neuronale Netze	Rule Induction	Genetic Programming	Case-based reasoning
Ergebnis aufwandsarm verfügbar	6%	1	1	3	3	9	9	9	9	3	1	1	3	3
Nachvollziehbar	3%	9	9	9	9	9	1	3	1	3	0	9	1	9
Hohe Schätzgenauigkeit und Treffsicherheit	15%	1	3	1	3	3	1	3	9	9	1	1	1	3
Akzeptanz durch oberes Management	15%	9	9	9	1	3	1	3	3	3				9
Akzeptanz durch Projektmitarbeiter	15%	3	3	3	3	3	3	3	9	9	1	1	1	9
Ohne Verwendung historischer Daten	0%	0	0	1	0	0	9	9	9	9	0	0	0	0
Kein Domänenwissen erforderlich	0%	3	3	9	3	1	0	0	0	0	9	3	9	3
leicht anpassbar (an neue Technologie, Entw.prozess,...)	8%	1	1	1	1	3	9	9	9	9	1	1	1	1
Ohne Zusatzkosten f. Softwarelizenz	0%	1	0	1	0	9	9	9	9	9	0	0	0	0
Kein statistisches Wissen erforderlich	0%	0	0	1	1	9	9	9	9	9	9	9	9	9
Automatisierbar	6%	9	9	9	9	3	0	1	0	1	9	9	9	9
Früh im Entwicklungsprozess verwendbar	20%	3	3	1	3	9	9	9	9	9	3	3	3	9
Leicht manipulierbar	0%	1	0	0	3	9	9	3	9	3	1	3	3	9
Objektiv (intersubjektiv durchführbar)	3%	9	9	9	9	0	0	0	1	3	9	9	9	3
Jederzeit wiederholbar m. möglicher Verbesserung	5%	1	1	1	1	9	9	9	9	9	1	1	1	9
Berücksichtigung einer Vielzahl v. Aufwandseinflüssen	5%	3	3	9	3	9	9	9	9	9	9	3	3	9
Unabhängig v. Test- o. Entwicklungsmethodik, -prozess, ...	0%	0	0	0	9	9	3	3	3	3	9	9	9	9
Absolute Bedeutung		3.9	4.2	3.9	3.1	5.2	4.7	5.4	7.2	7.0	2.3	2.2	2.2	7.0
Relative Bedeutung		6.7	7.2	6.7	5.3	9.0	8.1	9.4	12.3	12.0	4.0	3.8	3.7	11.9
Rangfolge		9	7	8	10	5	6	4	1	2	11	12	13	3

Bild 1: Matrix zur Auswahl geeigneter Aufwandsschätzmethoden

**Schritt 5: Aktualisierung des Vorgehens (bestehend aus lokalem Modell und ausgewählten, integrierten Schätzmethoden)**

Sowohl die Auswahl der Schätzmethoden als auch die Bestimmung der lokal wichtigsten Einflussfaktoren muss regelmäßig überprüft bzw. wiederholt werden, um Veränderungen am Entwicklungs- und Testprozess gerecht zu werden. Die Prozessänderungen in den Fallstudien (siehe folgenden Abschnitt) legten nahe, die relevanten Einflussfaktoren alle zwei Jahre zu überprüfen.

**5 Testaufwandsschätzung mithilfe der Methode TestASS**

Zur Schätzung des Testaufwands ist entscheidend, welches Ziel damit erreicht werden soll. Von den vier Zieldimensionen Quantität, d. h. funktionaler Umfang, Zeit, Kosten und Qualität wird mittels Test das Ziel Qualität direkt verfolgt, während Quantität, Zeit und Kosten Rahmenbedingungen des Tests darstellen. Daher basiert die Testaufwandsschätzung auf einem Vergleich des zu schätzenden Projekts mit abgeschlossenen Projekten, deren erreichtes Qualitätsziel dem angestrebten Qualitätsziel des zu schätzenden Projekts entspricht.

Das erreichte Qualitätsziel lässt sich messen, indem die nach Auslieferung gefundenen Fehler ausgewertet werden. Dabei können die Anzahl und Schwere der

Fehler in einem bestimmten Zeitraum als Vergleichskriterium herangezogen werden. Um dabei die Nutzungsintensität des Produkts zu berücksichtigen, die die Anzahl gefundener Fehler mitbestimmt, setzt man die gefundene Fehleranzahl beispielsweise ins Verhältnis zur Anzahl an Nutzern des Produkts.

Die mittels TestASS identifizierten relevanten Einflussfaktoren können in Verbindung mit verschiedenen Schätzmethode zur Aufwandsschätzung verwendet werden. Beispielsweise ermittelt man die Ausprägungen der Faktoren für das zu schätzende Projekt und setzt sie in die spezifizierte Modellgleichung ein. Das Modell in nicht-parametrisierter Form, d. h. die Faktoren ohne Gewichtung können zur Auswahl der Analogieobjekte im Rahmen der Methode Case-based reasoning oder zur Konfiguration eines künstlichen Neuronales Netzes herangezogen werden. Bei der Schätzung durch Experten (einzeln oder in einer Gruppe, informell oder formalisiert) werden die ermittelten Einflussfaktoren zur Berücksichtigung vorgegeben.

## 6 Validierung der Ergebnisse

Bisher wurde die Methode in zwei Organisationen eingesetzt, die betriebswirtschaftliche Standardsoftware herstellen. Dabei wurden jeweils lokale Modelle zur Testaufwandsschätzung entwickelt und validiert.

Ergebnisse in Organisation 1:

In dieser Organisation legten die Projektdaten nahe, separate Modelle für Neuentwicklungs- und Verbesserungsprojekte abzuleiten, was jedoch erfordert, dass je Projekttyp ausreichend viele Projekte analysiert werden. Die Analysen führten zu folgenden Modellgleichungen:

Modell für Neuentwicklungsprojekte:

$$TA = 17,1 - 6,0 * Abh + 6,7 * MuSpr + 1,0 Entw. erf. \text{ bzw.}$$

$TA = f$  (Abhängigkeit von Projektexternen, Anzahl Muttersprachen im Team, Erfahrung der Entwickler mit ihren Aufgaben, S)

Modell für Verbesserungsprojekte:

$$TA = 30,1 + 0,4 * ext. SS - 2,3 * Entw. erf. + 0,6 * Zeit \text{ bzw.}$$

$TA = f$  (Anzahl projektexterner Schnittstellen, Erfahrung d. Entwickler mit ihren Aufgaben, Zeitknappheit, S)

Dabei bezeichnet TA den Testaufwand relativ zum Entwicklungsaufwand und S die Störvariable. S enthält die Faktoren, die den Testaufwand beeinflussen, jedoch nicht als Funktionsvariable berücksichtigt sind (weil sie der Messung unzugänglich sind wie bspw. die Fähigkeiten der Entwickler).

Die Expertenrunde zur Validierung bestätigte den Einfluss der enthaltenen Variablen und beurteilte den jeweiligen Standardfehler (5,17 bzw. 7,07 für Neuent-

wicklungs- bzw. Verbesserungsprojekte) als annehmbar. Bei der Validierung anhand von weiteren Projektdaten erwies sich, dass der Testaufwand für Neuentwicklungsprojekte mittels Modell um 5 bzw. knapp 7 % zu hoch geschätzt wurde, was angesichts des geringen Aufwands durchaus akzeptabel erscheint. Für Verbesserungsprojekte ergaben sich deutlich höhere Abweichungen, weshalb dieses Modell anhand der zusätzlichen Daten anzupassen ist. In beiden Pilotprojekten wich der tatsächliche stark vom geschätzten Testaufwand ab. Allerdings konnte die Erreichung des Qualitätsziels anhand der Fehler, die nach Auslieferung gemeldet werden, nicht gemessen werden, sodass keine Aussage möglich ist, ob der überschätzte Testaufwand ausreichend war oder ein der Schätzung entsprechender Testaufwand zielführender gewesen wäre.

Ergebnisse in Organisation 2:

Die Regressionsgleichung, die gemäß den Gütekriterien das beste Modell des Testaufwands in dieser Organisation darstellt, enthielt die Faktoren „Bekanntheit der eingesetzten Entwicklungstechnologien“ und „Anzahl externer Schnittstellen“. Allerdings wies der errechnete Standardfehler der Gleichung einen sehr hohen Wert auf, d. h. die Gleichung bildete die modellierten Daten sehr ungenau ab, sodass sie zur Testaufwandsschätzung ungeeignet erschien. Dies lag zum einen am relativ geringen Umfang der Datenbasis (16 Datensätze), zum anderen an nicht berücksichtigten Faktoren, die entweder schwierig zu operationalisieren waren (z. B. „Zusammenarbeit mit dem Auftraggeber“) oder nur in einzelnen Fällen relevant und daher nicht modelliert waren (z. B. „gegenüber der Planung reduzierte finanzielle Mittel und Mitarbeiteranzahl für den Test“). Um dennoch ein anwendbares Modell zur Testaufwandsschätzung abzuleiten, sind in diesem Fall weitere Datensätze sowie die Operationalisierung der Faktoren erforderlich, die sich anhand von Interviewaussagen in der Einzelfallanalyse als relevant zeigten, jedoch bei der Datensammlung nicht berücksichtigt und nicht für alle Projekte erhoben worden waren.

## 7 Ausblick

Die Ergebnisse der mit der Methode TestASS erstellten lokalen Testaufwandsschätzmodelle sind nicht eindeutig, und ein direkter Vergleich mit anderen Methoden zur Testaufwandsschätzung steht noch aus. Dennoch ist der Ansatz viel versprechend, da er sich an organisationspezifischen Besonderheiten orientiert und als Basis quantitative mit qualitativen empirischen Daten verbindet. Die Methode erhebt nicht den Anspruch, ein allgemein gültiges, d. h. in jeder Organisation unabhängig vom verfolgten Entwicklungsprozess einsetzbares Modell zur Testaufwandsschätzung bereitzustellen. Vielmehr muss der jeweils verfolgte Entwicklungsprozess nicht vermessen, sondern lediglich in allen untersuchten und zu schätzenden Projekten gleichermaßen eingehalten werden, damit das Mo-

dell zur Testaufwandsschätzung verwendet werden kann. Daher zeichnet sich die Methode durch hohe Flexibilität aus.

Bei Anwendung der Methode stellt sich erwartungsgemäß weiterer Nutzen ein:

- Die Ergebnisse der Aufwandsschätzung sind dank gemeinsamer Datenbasis organisationsweit vergleichbar.
- Auch ohne Schätzerfahrung lässt sich der Aufwand systematisch und nachvollziehbar schätzen.
- Mit jeder Erweiterung der Projektdatenbasis können genauere Schätzungen erzielt werden.
- Zeitliche, personelle und finanzielle Restriktionen können als Einflussfaktoren modelliert werden.
- Insbesondere die mittels des geschätzten Testaufwands erreichbare Produktqualität wird bei der Schätzung des Testaufwands berücksichtigt.

### **Literaturhinweise**

1. Basili, Victor R.; Rombach, H. D. (1988), The TAME project: Towards improvement-oriented software environments. In: IEEE Transactions on Software Engineering, Vol. 14, 1988, Nr. 6, S. 758-773
2. Boehm, Barry W., Abts, C.; Horowitz, E.; Madachy, R.; Reifer, D.; Clark, B.K.; Steece, B.; Brown, A.W.; Chulani, S. (2000), Software cost estimation with Cocomo II. Upper Saddle River 2000
3. Cangussu, Joao W.; DeCarlo, Raymond A.; Mathur, Aditya P (2002), A formal model of the software test process. In: IEEE Transactions on Software Engineering, Vol. 28, 2002, Nr. 8, S. 782-796
4. Grady, Robert B. (1992), Practical Software Metrics for Project Management and Process Improvement. Upper Saddle River 1992
5. Kadoda, Gada; Cartwright, Michelle; Chen, Liguang; Shepperd, Martin (2000), Experiences using case-based reasoning to predict software project effort. ESERG Technical Report No. 00-09, Bournemouth University, 2000.
6. Ludewig, Jochen; Lichter, Horst (2007), Software Engineering: Grundlagen, Menschen, Prozesse, Techniken. Heidelberg 2007
7. Myrtveit, Ingunn, Stensrud, Erik, Olsson, Ulf (2001), Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods, in: IEEE Transactions on Software Engineering, Vol. 27, 2001, Nr. 11, S. 999-1013
8. Nageswaran, Suresh (2001), Test effort estimation using Use Case points. In: Proceedings of the Quality Week 2001, San Francisco, 2001

9. Sneed, Harry; Jungmayr, Stefan (2006), Produkt- und Prozessmetriken für den Softwaretest. In: Informatik Spektrum, Jg. 29, 2006, Nr. 1, S. 23-39
10. Sneed, Harry; Baumgartner, Manfred; Seidl, Richard (2007), Der Systemtest. München / Wien 2007
11. Stevens, J. (1986), Applied Multivariate Analysis for the Social Sciences. Hillsdale (New Jersey, USA) 1986
12. Takahashi, Muneo; Kamayachi, Yuji (1989), An Empirical Study of a Model for Program Error Prediction. In: IEEE Transactions on Software Engineering, Vol. 15, 1989, Nr. 1, S. 82-86
13. Thaller, Georg Erwin (2000a), Software-Test. Verifikation und Validation. Hannover 2000
14. Thaller, Georg E. (2000b), Software-Metriken einsetzen, bewerten, messen. 2. Aufl., Berlin 2000
15. Trauboth, Heinz (1996), Software-Qualitätssicherung. Konstruktive und analytische Maßnahmen. In: Endres, A.; Krallmann, H.; Schnupp, P. (Hrsg.): Handbuch der Informatik. Band 5.2. 2. Aufl., München / Wien 1996