

Warum Prüfen 50-mal länger dauert als Lesen

... und andere Überraschungen aus der Welt der Software-Reviews

Zusammenfassung:

Bei der Diskussion über Software-Reviews werden einige Zahlenangaben, die von Fachleuten genannt werden,

- von den meisten Softwareentwicklern intuitiv in ganz anderer Größenordnung eingeschätzt
- und daher stark angezweifelt.



Peter Rösler, München
Reviewtechnik-Trainer

Zwei dieser Zahlenangaben, die den meisten Software-Entwicklern besonders unplausibel vorkommen, sind:

- Die optimale Inspektionsrate für Textdokumente beträgt nur ca. **1 Seite pro Stunde** (und liegt damit um ca. den **Faktor 50** unter der reinen Lesegeschwindigkeit).
- Das durchschnittliche Review findet nur ca. **4-7%** der im Dokument vorhandenen Fehler.

Zusammenfassung (3)

3

- Während des Vortrags können die Zuhörer an einigen Kurzexperimenten und Schätzungen teilnehmen, deren Ergebnisse die obigen Zahlenangaben plausibel machen.
- An den Kurzexperimenten und Schätzungen beteiligten sich bisher ca. 700 Teilnehmer von 59 Reviewtechnik-Seminaren und -Vorträgen im Zeitraum von Oktober 2004 bis August 2008.

1. Einführung

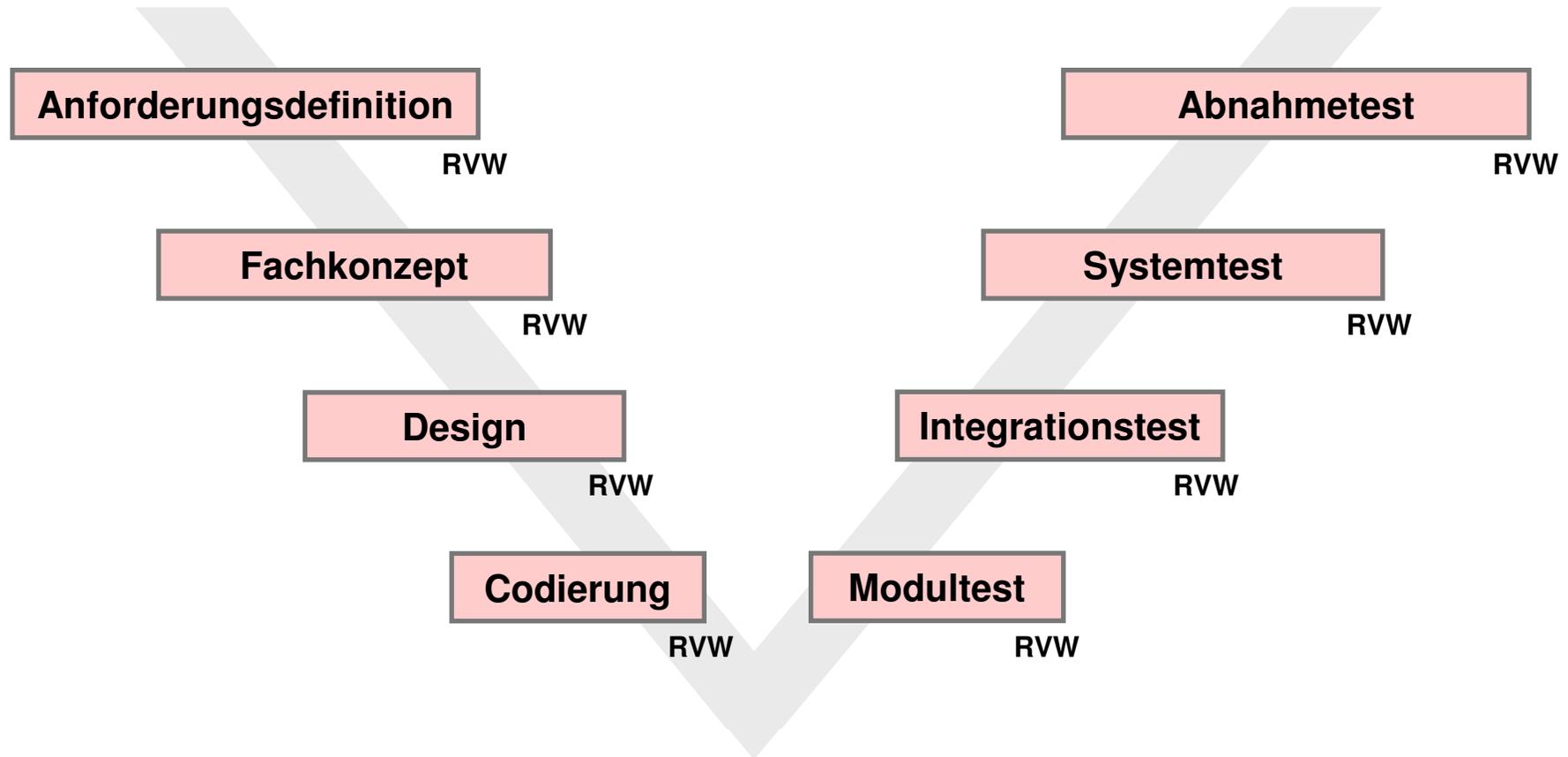
4

- Die in diesem Beitrag betrachteten Software-Reviews werden oft auch als „Software-Inspektionen“, „Fagan/Gilb style inspections“ oder „Peer Reviews“ bezeichnet.
- Definitionen:
“major defect” (im Gegensatz zu “minor defect”) : Fehler, der möglicherweise erheblich höhere Kosten verursacht, wenn er später gefunden wird als jetzt.

1 Seite = 300 Wörter [3]

Reviews werden in jeder Projektphase eingesetzt

5



Types of Review of Documents

6

Walkthrough: [Presentation Review] Activity: understanding

- author guides the group through a document and his or her thought processes, so all understand the same thing, consensus on change

Review: [Management-Review/Projektstatus-Review] Activity: decision making

- group discusses document and makes a decision about the content, e.g. how something should be done, go or no-go decision

Inspection: Main activity: find defects

- formal individual and group checking, using sources and standards, according to detailed and specific rules



Rollen der Teilnehmer

7

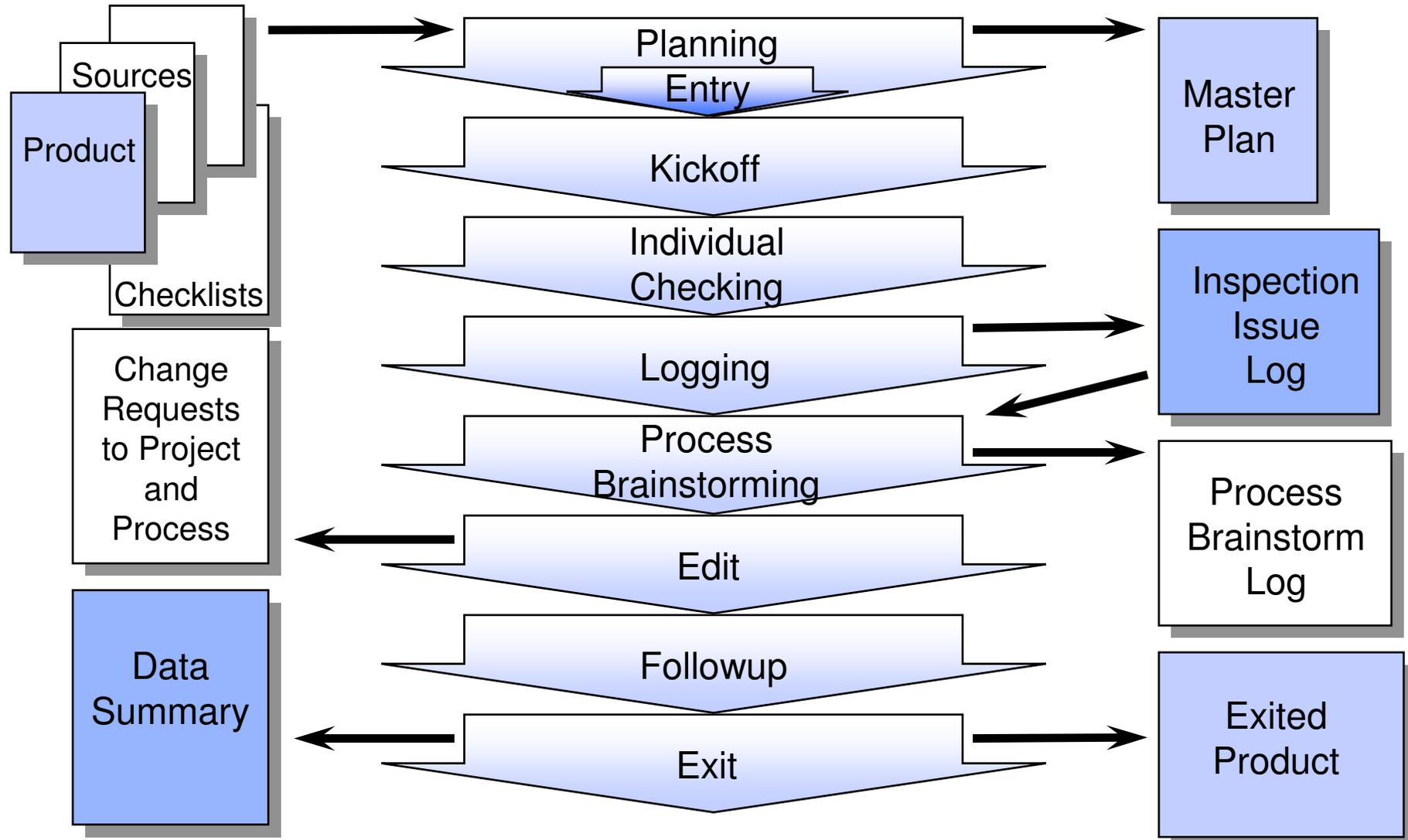
- Moderator
- Autor
- Protokollführer
- Reviewer
- Vorleser/Reader (nur wenn „double checking“ gemacht wird)



Ein Teilnehmer kann mehrere Rollen übernehmen.
Einzige Einschränkung: der Autor darf zusätzlich höchstens die Rolle eines Reviewers übernehmen.

Overall Process Map

8



Phase „Individual Checking“

9

- Von den Fehlern, die das Reviewteam insgesamt entdeckt, werden **80%** im „Individual Checking“ gefunden und **20%** im „Logging meeting“, sofern dort mit „Double Checking“ erneut Fehlersuche betrieben wird. [1]
- Da in den meisten durchgeführten Reviews auf „Double Checking“ verzichtet wird, werden in der Praxis eher **95%** der durch das Reviewteam entdeckten Fehler im „Individual Checking“ gefunden. [2]



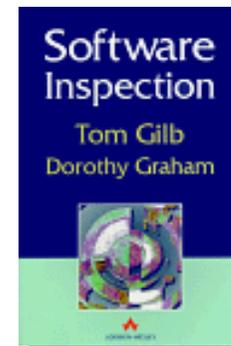
2. Lese- und Prüfgeschwindigkeiten im Vergleich

2.1 Die optimale Inspektionsrate

10

- In der Phase „Individual Checking“ wendet ein Reviewer alle ihm zur Verfügung stehenden sinnvollen Prüfstrategien an.
- Für Textdokumente dauert das erfahrungsgemäß **1 Seite pro Stunde**. ([1], Gilb/Graham 1993)
- Bandbreiten für die optimale Inspektionsrate: **$1 \pm 0,8$ Seiten pro Stunde** [3]

„Software Inspection“ von Tom Gilb
und Dorothy Graham, 1993

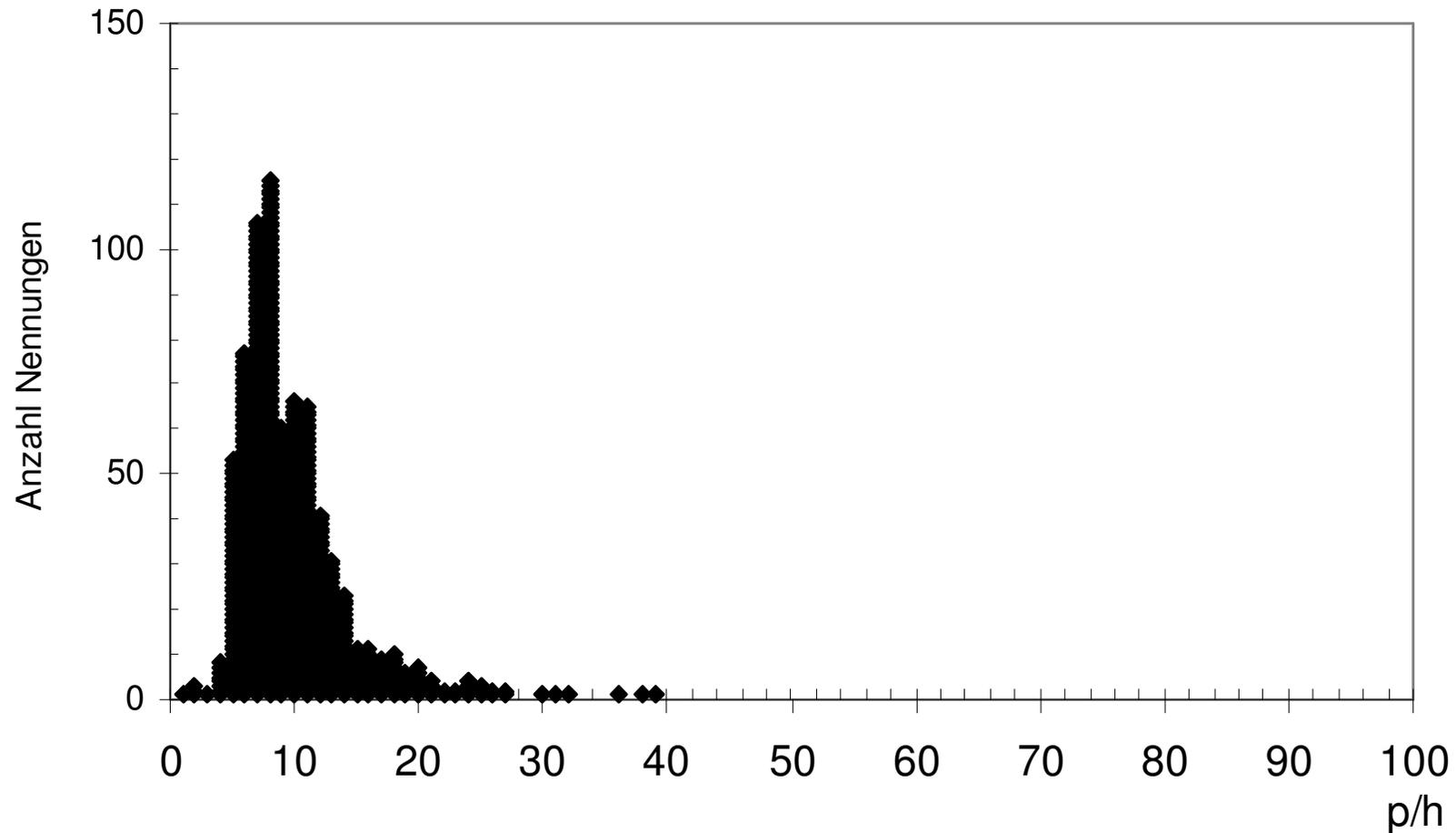


- Mit folgenden Kurzexperimenten und Schätzungen kann der unplausibel klingende Wert „1 Seite pro Stunde“ nachvollzogen werden:
- In einem ersten Kurzexperiment prüfen die Teilnehmer einen typischen Spezifikationstext auf „minor defects“.

Spezifikationstext: „...For each flight in the database, this configuration file will be examined line by line. The assignment part of the first rule for which the condition part matches the carrier and trip number will be used as the new load control value. ...”).

„minor defects“-Inspektionsraten

12



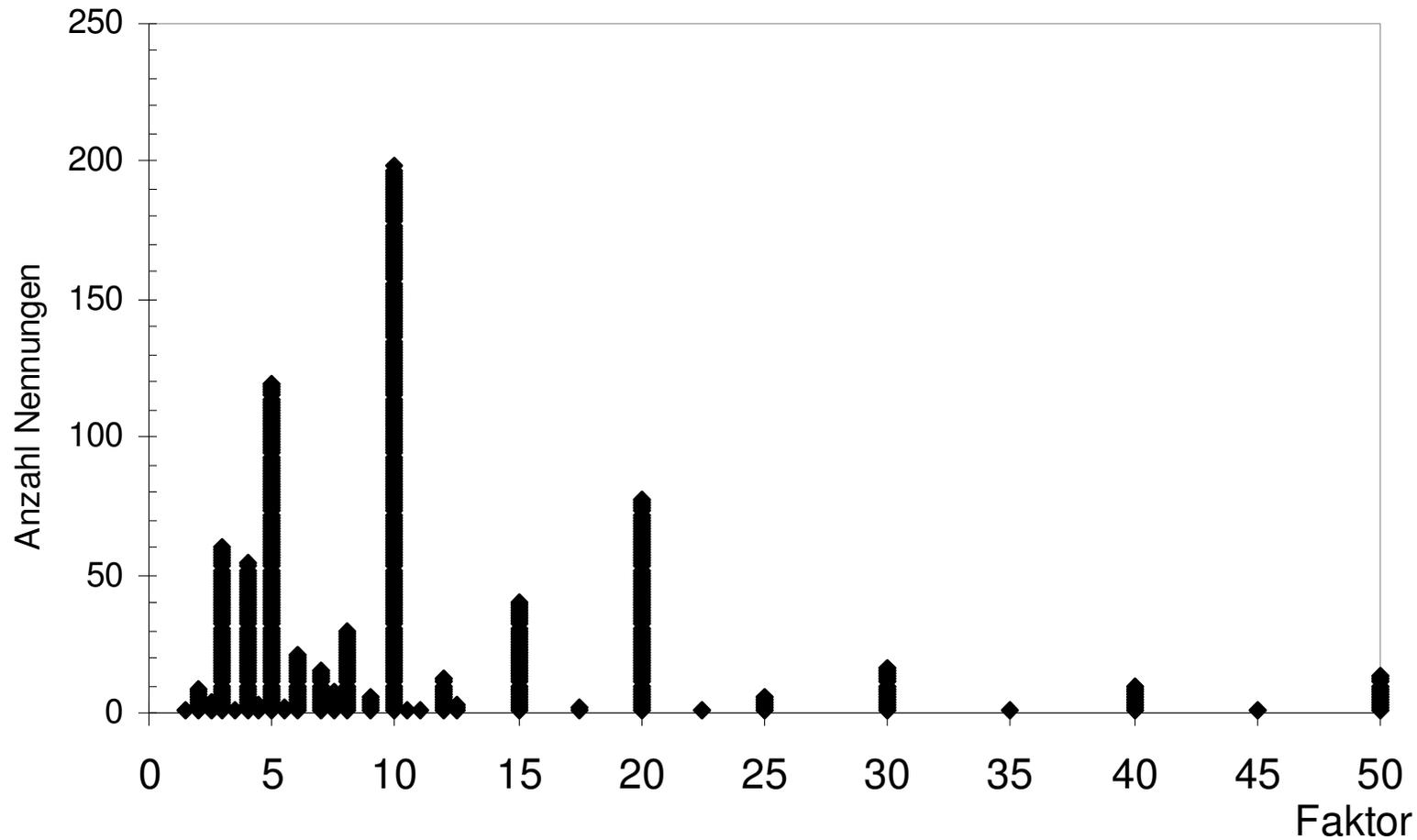
Mittelwert: 9,8 p/h

Anzahl Datenpunkte: 729



- Nach dem Kurzexperiment mit der „minor defects“-Inspektionsrate werden die Teilnehmer gefragt:
„Wie viel mal mehr Zeit (zusätzlich) würden Sie schätzungsweise benötigen, wenn Sie auch die inhaltlichen Fehler, also die **Major defects**, finden wollten?“

Geschätzte Zusatzzeit für „Major defects“-Suche



Mittelwert: Faktor 14,0

Anzahl Datenpunkte: 735



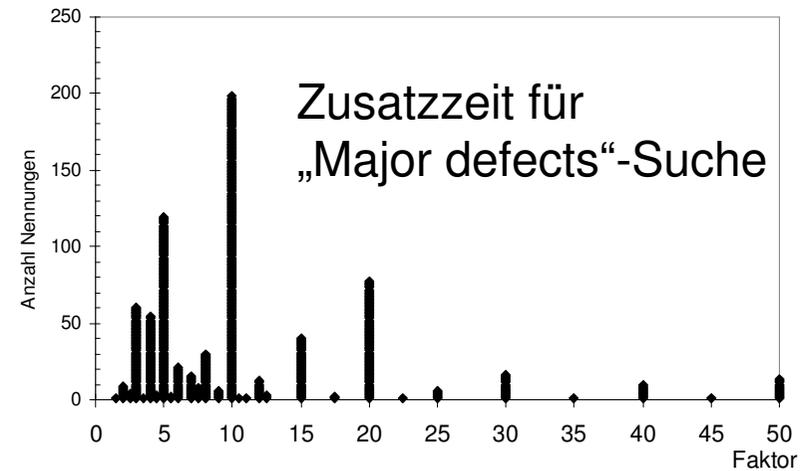
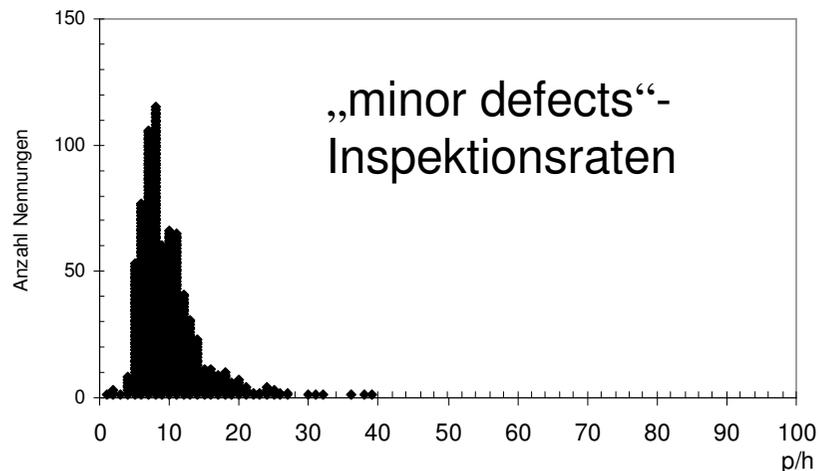
Die optimale Inspektionsrate

- Aus dem Kurzexperiment und der Schätzung ergibt sich folgende optimale Inspektionsrate:

optimale Inspektionsrate =

„minor defects“-Inspektionsrate / (1+Zusatzzeitfaktor) =

$$9,8 \text{ p/h} / (1+14,0) = \mathbf{0,7 \text{ p/h}}$$

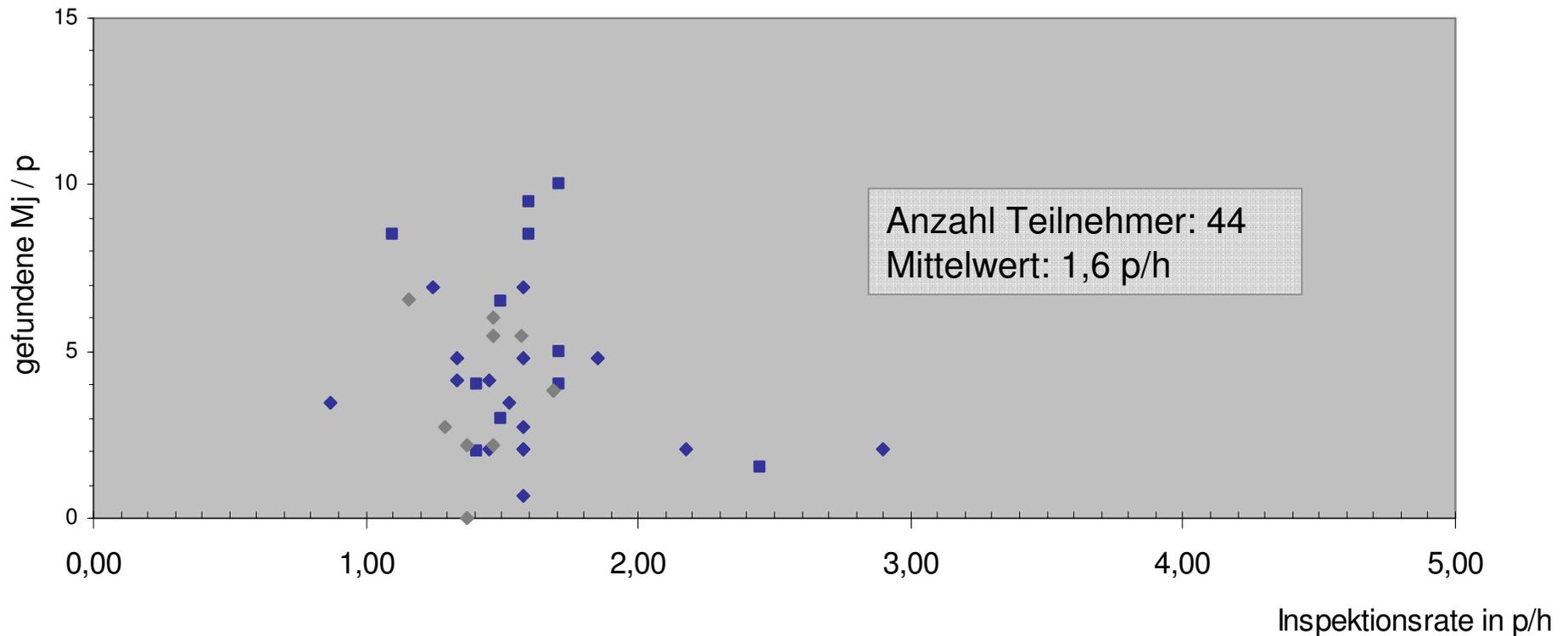


Selber ausprobieren ist noch überzeugender

16

Komplette Übungsinspektion auf Textdokumenten

◆ Spielregel 1,45 p ■ Spielregel 2,0 p ◆ Systemarchitektur 1,83 p



Optische Täuschungen

17



Die Diagonalen sind parallel zueinander.

Wer's nicht glaubt, kann jederzeit nachmessen. Bei diesem Bild darf man seiner Intuition nicht trauen.

Die optimale Inspektionsrate beträgt ca. 1 Seite pro Stunde.

Wer's nicht glaubt, kann jederzeit nachmessen! Auch hier darf man seiner Intuition nicht trauen!

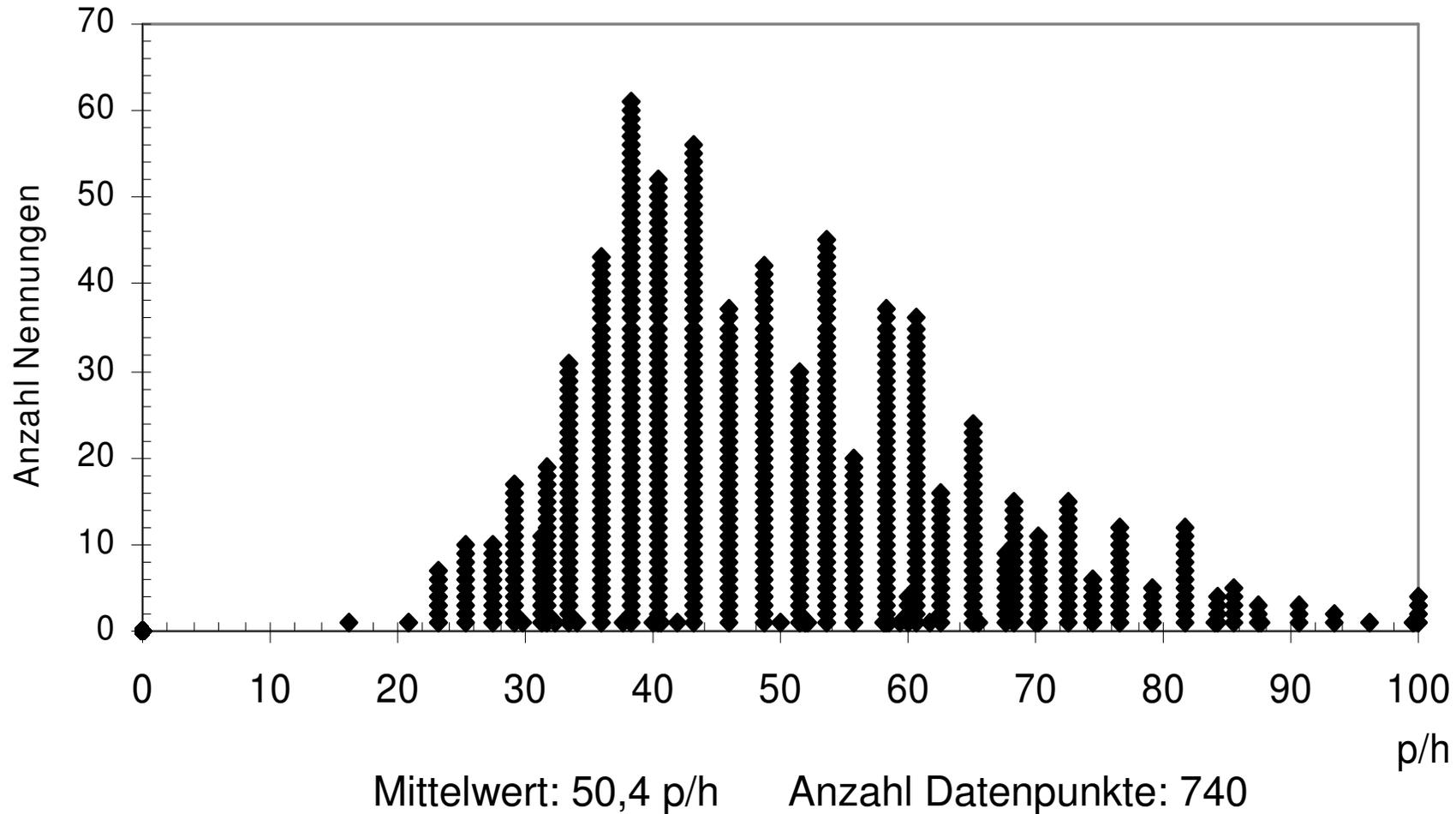


2.2 Warum Prüfen 50-mal länger dauert als Lesen

- Das „Naturtempo“ beim Lesen von (deutschsprachigen) Texten beträgt ca. 240 Wörter pro Minute. [4]
- Das sind umgerechnet 48 Seiten pro Stunde.
- In einem Kurzexperiment ermitteln die Teilnehmer ihre eigene Lesegeschwindigkeit an einem Übungstext.

Textausschnitt: „...Ein skeptischer Autor beschloss, Magliabechis Bekanntheitsgrad wegen seines Schnelllesens und Gedächtnisses auf die Probe zu stellen und gab ihm ein neues Manuskript, das er vorher nie gesehen haben konnte. ...“ [5])

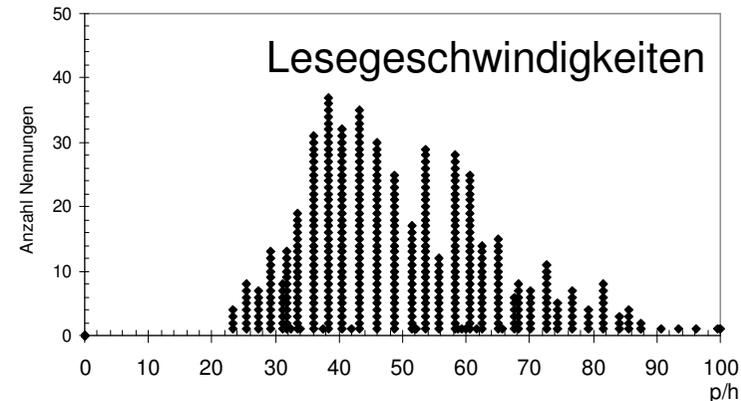
Lesegeschwindigkeiten



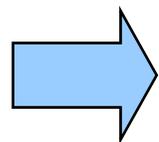
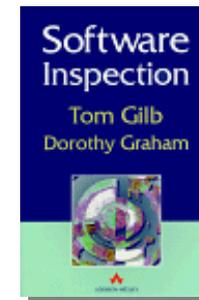
Prüfen dauert 50 mal länger als Lesen

20

- Lesegeschwindigkeit:
ca. 50 Seiten pro Stunde



- optimale Inspektionsrate:
ca. 1 Seite pro Stunde



„Man kann zwar ca. 50 Seiten pro Stunde lesen,
aber nur ca. 1 Seite pro Stunde gründlich prüfen.“

3. Effektivität von Reviews

21

- Das Reviewteam findet ca. **50%** der im Dokument vorhandenen Fehler [1][2], wenn die optimale Inspektionsrate eingehalten wird, ansonsten:
- Tom Gilb / Kai Gilb: “We have found Inspection to operate in the area of about **3%** (**the normal for inspections not optimized with optimal checking rates etc.**) to about 88%.” [6]
- Mit den im Folgenden gezeigten Kurzexperimenten und Schätzungen kommt man auf einen ähnlichen Wert: nur ca. **4-7%** der Fehler werden in typischen Reviews entdeckt.



3.1 Typische Inspektionsraten beim Review eines Textdokuments

22

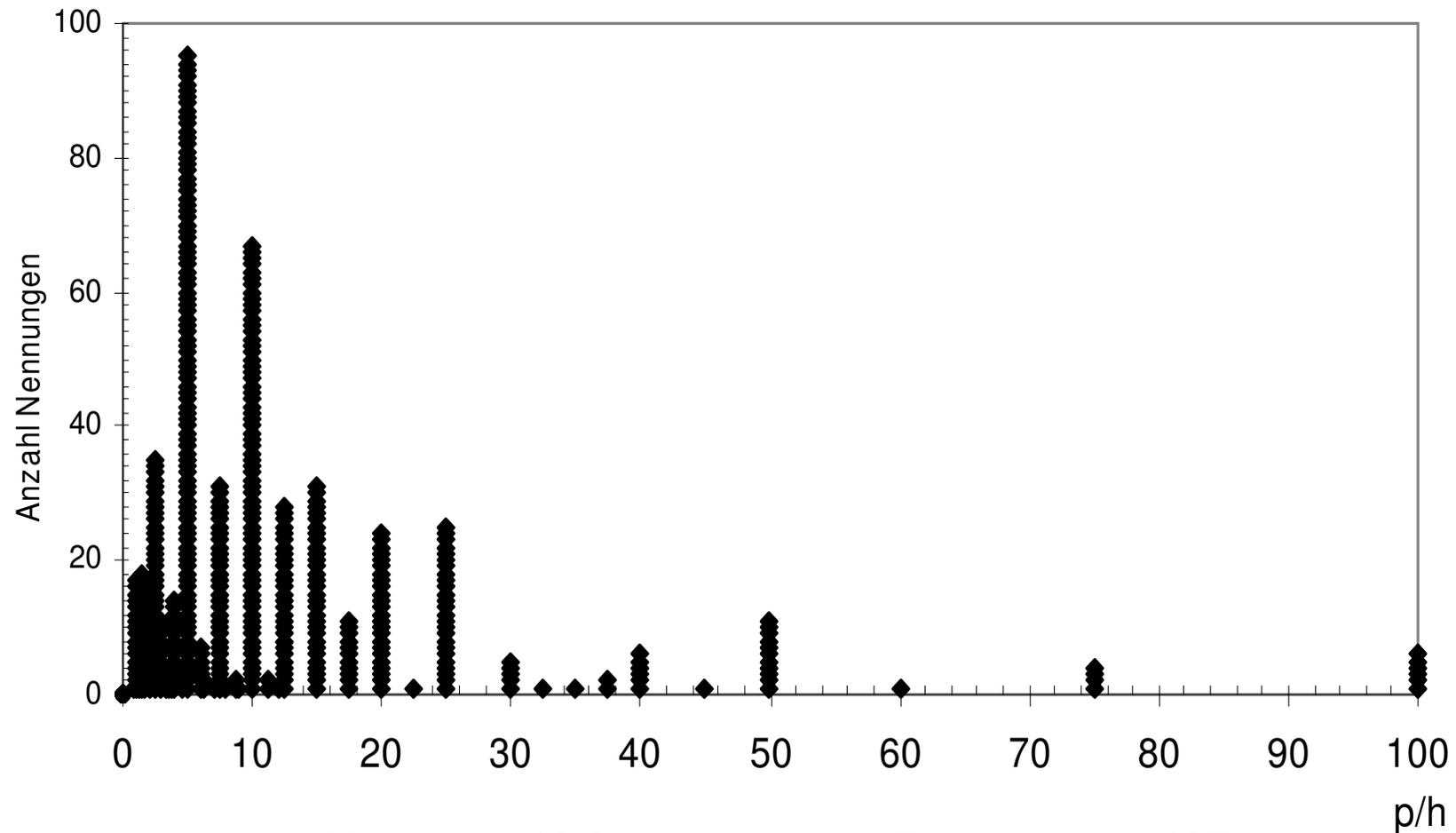
- In einer Umfrage gaben die Teilnehmer an, mit welchen Inspektionsraten die Reviews in ihrer Firma typischerweise durchgeführt werden.

„In Ihrer Firma soll ein Review eines Textdokuments stattfinden. Das Dokument ist so groß, dass es in einzelnen Paketen geprüft wird, jedes Paket in einer eigenen Reviewsitzung. Angenommen, jeder Reviewer kann ca. 2 Stunden Aufwand in die Vorbereitung für eine Reviewsitzung investieren. Versuchen Sie abzuschätzen, aus wie vielen Seiten ein solches Paket in Ihrer Firma typischerweise besteht! ...“



Typische Inspektionsraten beim Review eines Textdokuments

23



Mittelwert: 13,3 p/h
Median: 7,5 p/h

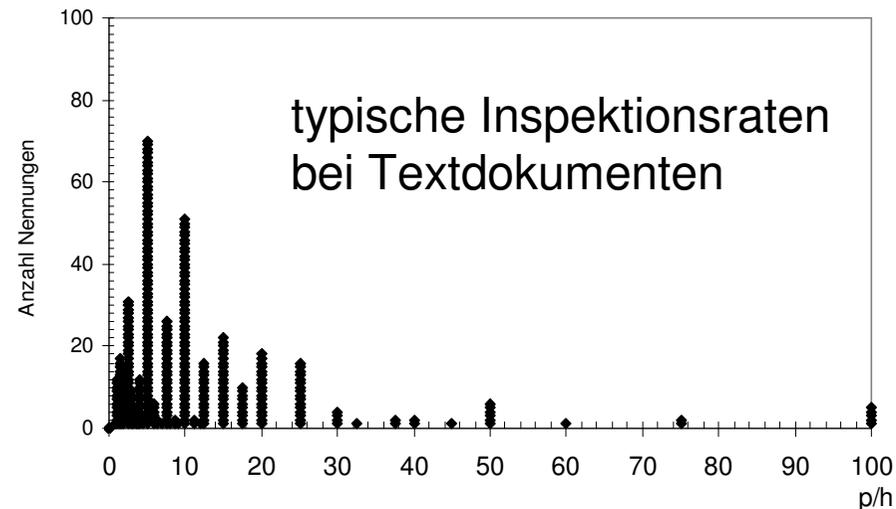
Anzahl Datenpunkte: 497



Typische Reviews: um Faktor 13 zu schnell!

- Die optimale Inspektionsrate (1 Seite pro Stunde) wird in typischen Reviews um den Faktor 13 verfehlt.

(Bzw. um Faktor 7,5, wenn man mit dem Medianwert rechnet)

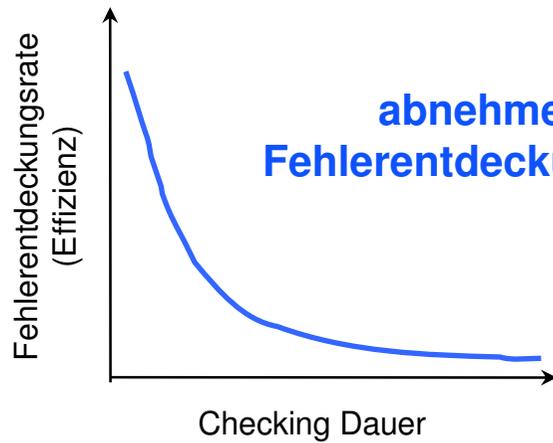


- Die Werte aus der Umfrage entsprechen den publizierten Erfahrungen:

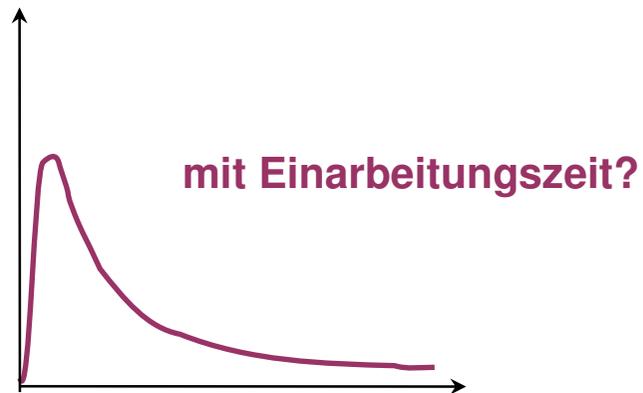
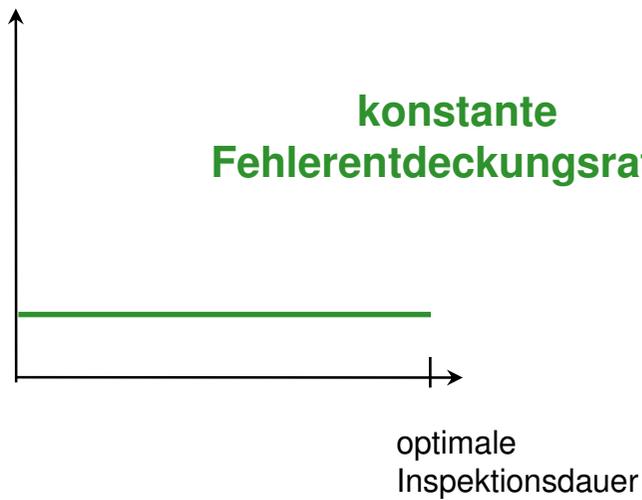
„we typically review documents at **five** to **twenty** pages per hour, when systematic calculation and application of optimum rates are not applied.” [1]

Wann findet ein Reviewer die Fehler in der Phase „Individual Checking“?

25



konstante Fehlerentdeckungsrate?

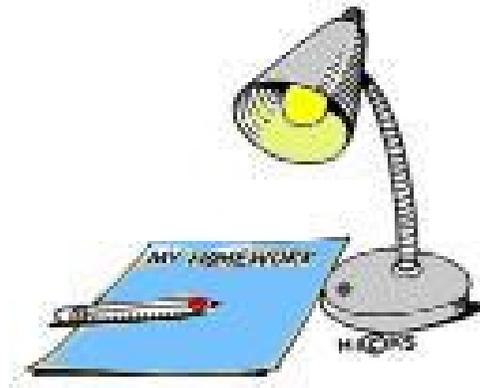


mit Einarbeitungszeit?



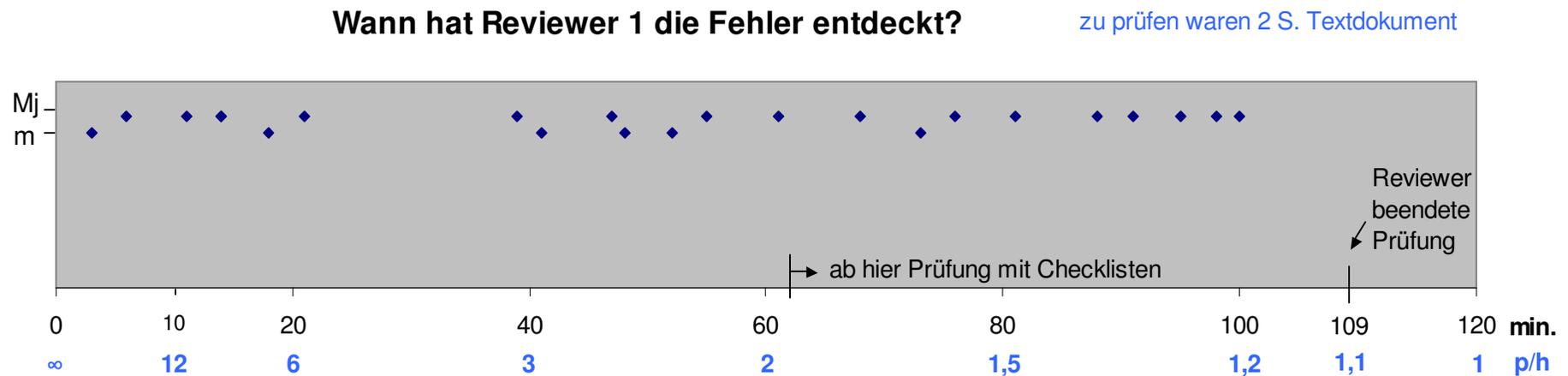
3.2 Hypothese: Im „Individual Checking“ werden die Fehler zeitlich ungefähr gleichverteilt gefunden

- 28 Reviewer wurden gebeten, bei jedem entdeckten Fehler den Zeitpunkt minutengenau zu protokollieren.



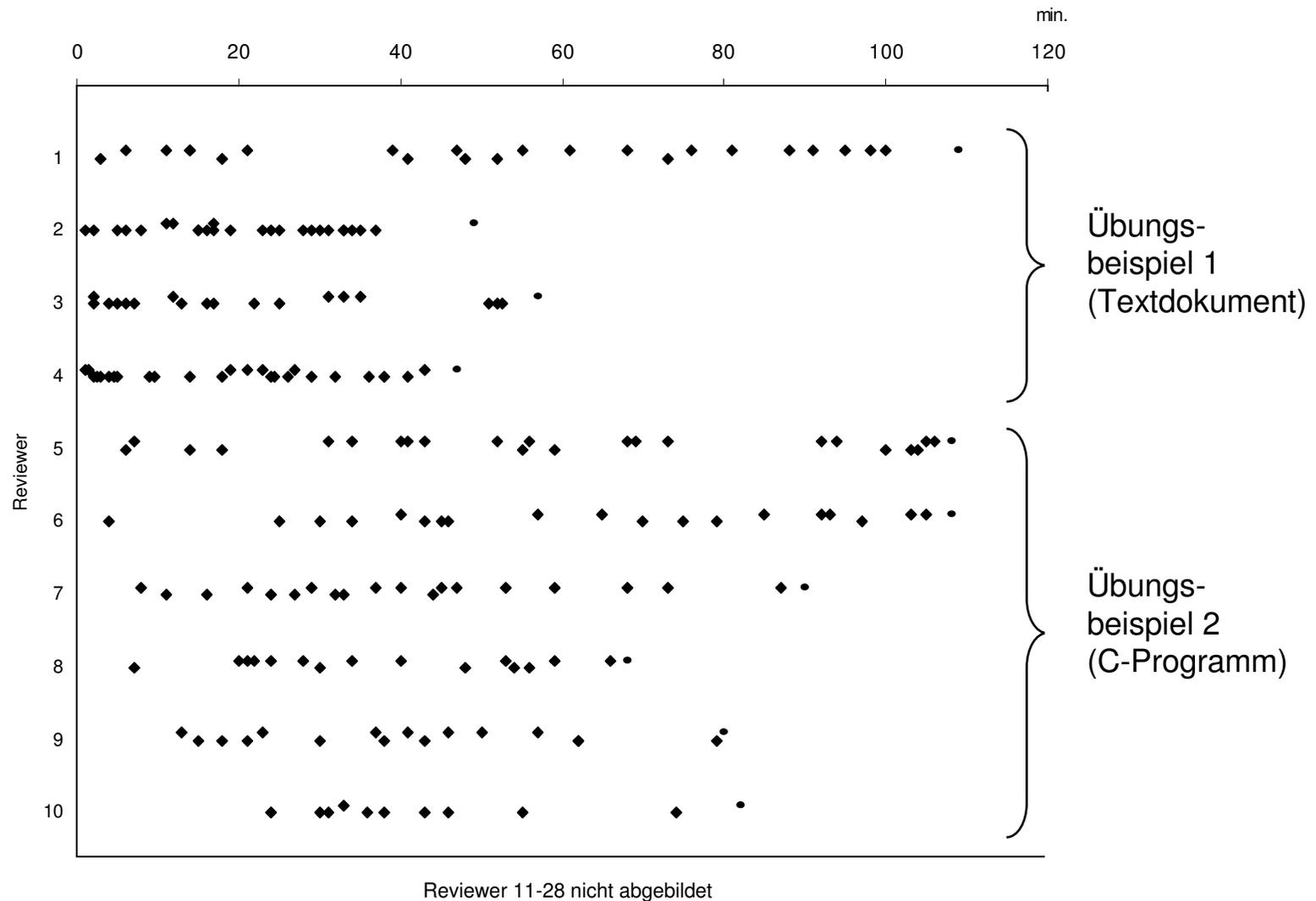
Zeitpunkt der Fehlerentdeckung in Phase „Individual Checking“, Reviewer 1

27

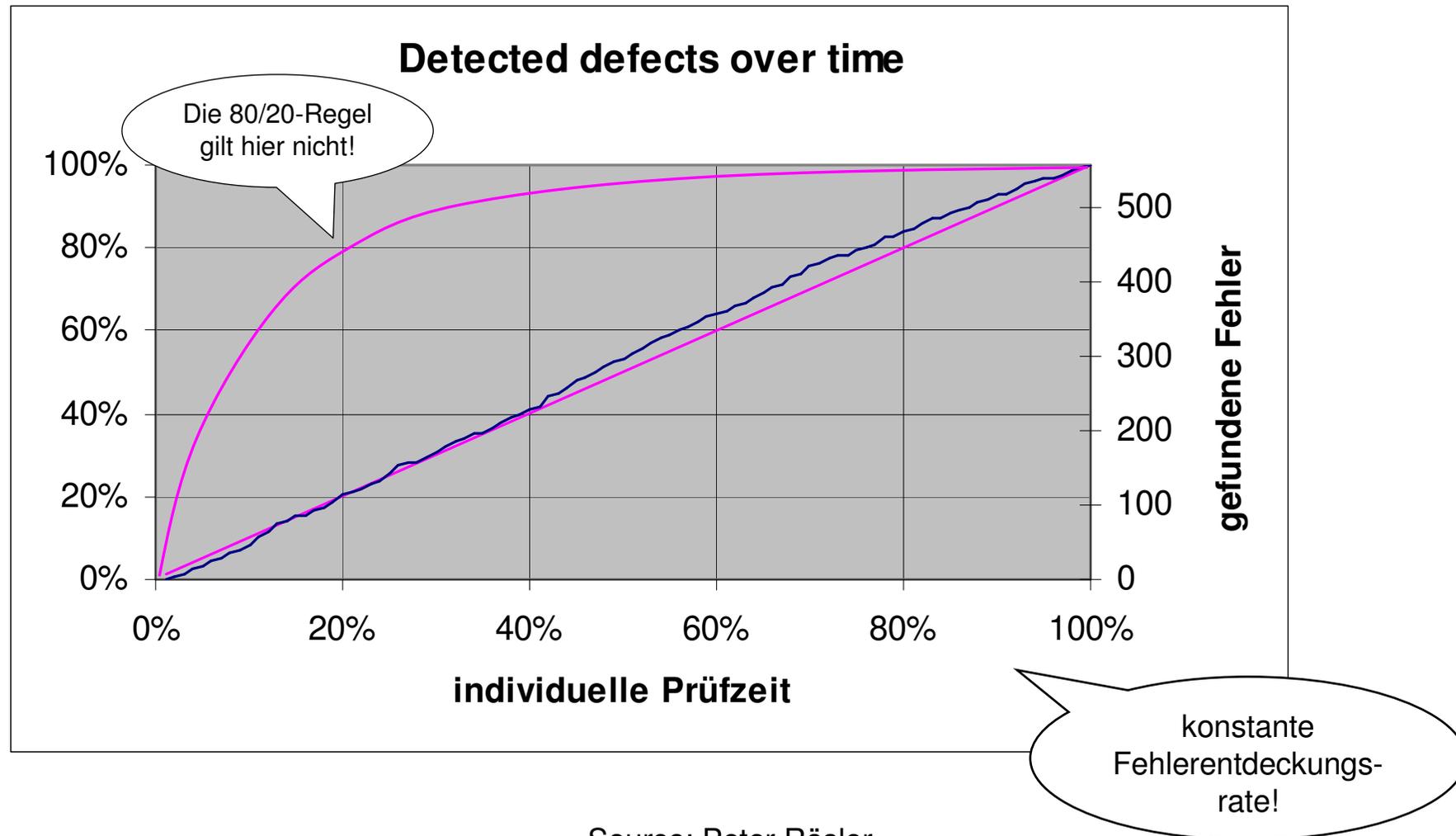


Zeitpunkt der Fehlerentdeckung in Phase „Individual Checking“

28

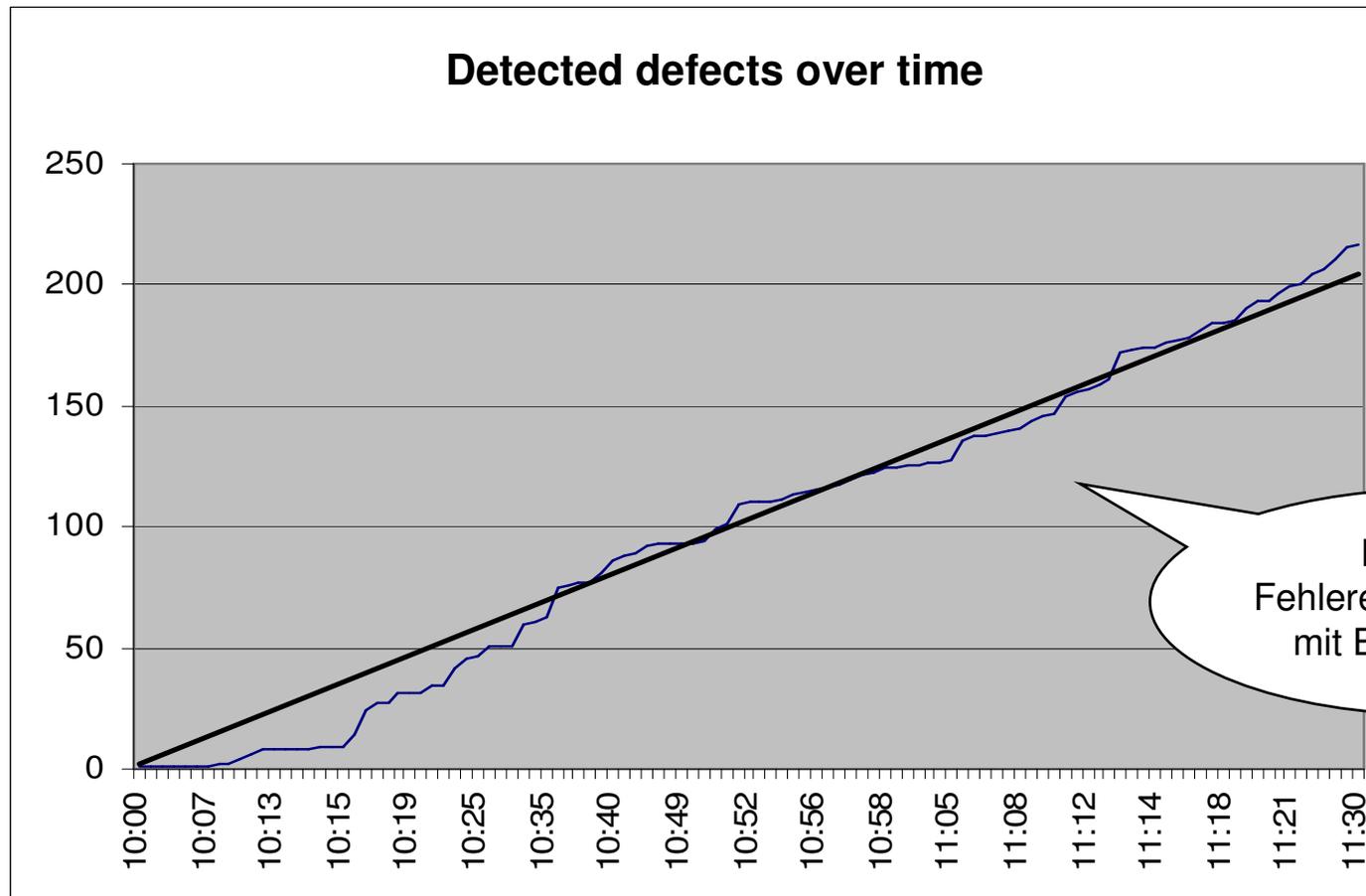


Kumuliertes Histogramm für 28 Reviewer



Ähnlicher Befund beim Review von UML-Modellen

30



Source: Górski / Jarzębowicz: Development and validation of a HAZOP-based inspection of UML models, in: Proceedings of the 3rd World Congress for Software Quality, Erlangen, 2005

Lange bekannter indirekter Hinweis: „Defect Density against Inspection Rate“

31

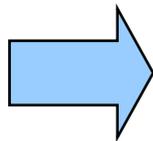
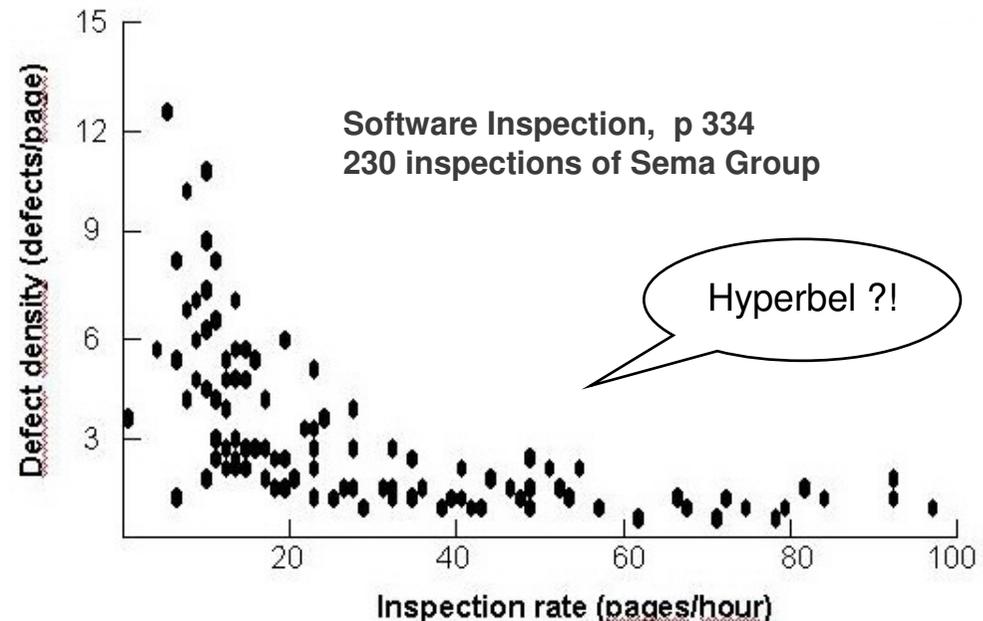
Uwe Schöffel, Softlab 1998:

“Die Anzahl der Fehler, die man pro Stunde findet, scheint laut dieser Kurve konstant zu sein.”

Niels Malotaux:

“... this figure seems to indicate that
Inspection rate * Defect density = constant

The implication is that the total number of defects found per hour is constant and does not depend on the number of pages inspected in that hour.“



Die Effizienz (Mj/h) ist konstant (für ein gegebenes Dokument)



3.3 Schlussfolgerung: das durchschnittliche Review findet nur ca. 4-7% der vorhandenen Fehler

- Aus der konstanten Fehlerentdeckungsrate folgt:
„Das Verfehlen der optimalen Inspektionsrate um Faktor x bewirkt ein Sinken der Effektivität um Faktor x “ ($x \geq 1$).
- Da in den realen SW-Projekten ca. 13 Seiten pro Stunde geprüft werden und damit die optimale Inspektionsrate um ca. Faktor 13 verfehlt wird, werden diese Reviews anstelle der möglichen 50% **nur ca. 4%** der vorhandenen Fehler entdecken! (Bzw. **nur ca. 7%**, wenn man mit dem Medianwert 7,5 rechnet)
- Diese niedrigen Werte gelten nur für Textdokument-Reviews! (Daten für Code-Reviews s. Notizansicht)

Ein Dokument mit 2 Methoden geprüft

33

Systemarchitektur mit 220 Seiten

- geprüft mit herkömmlichem Review (alle 220 Seiten)
- und geprüft mit Inspektion (nur 13-seitiges Kapitel)

	Anzahl Reviewer	Inspektionsrate in Phase „Individual Checking“	gefundene Mj defects im 13-seitigen Kapitel	Effektivität
herkömmliches Review	geplant: 23 real: 7	nichts geplant, real: ≥ 50 p/h	1	max. 2% (= $1/16 * 30\%$)
Inspektion	geplant: 4 real: 4	geplant: 4,3 p/h real: 4,1 p/h	16	max. 30% (geschätzt)

Ein Dokument mit 2 Methoden geprüft (2)

Die Effektivität bricht ungefähr um den Faktor ein, um den zu schnell inspiziert wird.



Die Effizienz hängt nicht von der Inspektionsrate ab.

(Sondern von der Fehlerdichte im Dokument, s. Notizansicht der Folie)



	Inspektionsrate	Effektivität	Aufwand	gefundene Mj defects	Effizienz
herkömmml. Review	≥ 50 p/h	max. 2%	≈ 22 h	11 (in 220 p)	$\approx 0,5$ Mj/h
Inspektion	4,1 p/h	max. 30%	≈ 35 h	16 (in 13 p)	$\approx 0,46$ Mj/h

Ein Dokument mit 2 Methoden geprüft (3)

35

Hochrechnung, wenn alle 220 Seiten mit Inspektionen geprüft worden wären:

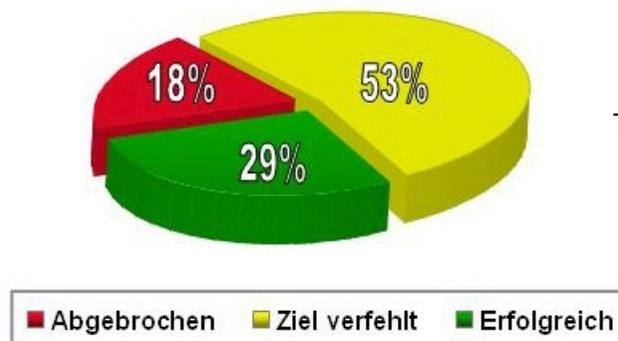
	Aufwand	gefundene Mj defects	Eingesparte Stunden *
herkömmliches Review	≈ 22 h	11 (in 220 p)	≈ 110 h (ROI ≈ 5:1)
Inspektionen	≈ 592 h	≈ 271 (in 220 p)	≈ 2660 h (ROI ≈ 4,5:1)

* Estimated time for find&fix a Major later: 12h



Schlussbemerkung

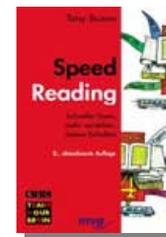
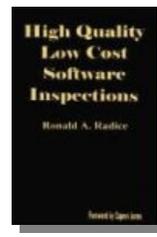
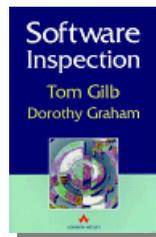
- Die niedrige Effektivität von 4-7% in Textdokument-Reviews legt nahe, dass die Qualitätssicherung in den meisten SW-Projekten kaum einen Einfluss auf den Projekterfolg hat:
- Ein Projekt, das scheitern würde, weil in den Dokumenten zu viele Fehler stecken, scheitert wahrscheinlich auch dann noch, wenn die „Qualitätssicherung“ mit ihren Reviews ca. 93-96% dieser Fehler übersieht!



Source:
The Standish Group,
s.a. Notizseite



1. Gilb, Tom / Graham, Dorothy: Software Inspection, Addison-Wesley, 1993,
2. Radice, Ronald A.: High Quality Low Cost Software Inspections, Paradoxicon Publishing, 2002
3. www.gilb.com (Download Center), "Optimizing Inspection" von Tom Gilb
4. Michelmann, Rotraut / Michelmann Walter U.: Effizient lesen, Gabler Verlag, Wiesbaden, 1. Auflage 1995
5. Buzan, Tony: Speed Reading, mvg-verlag, Landsberg-München, 2002
6. www.gilb.com, Topic "Insp., Quality Control", Thread "Completeness of candidate doc.", Posting von Kai Gilb vom 17.09.2004
7. Gorski, J. / Jarzebowicz A.: Development and validation of a HAZOP-based inspection of UML models, in: Proc. of the 3rd World Congress for Software Quality, Erlangen, 2005



Diskussion

5

Reviews werden in jeder Projektphase eingesetzt

Quelle: [unintelligible]

6

Types of Review of Documents

Walkthrough - Activity: understanding
 * author guides the group through a document and lets or her thought processes, to all understand the same thing, consensus on change.

Desktop - Activity: decision making
 * group discusses document and makes a decision about the content, e.g. how something should be done, go or no-go decision

Inspection - Main activity: find defects
 * formal individual and group checking, using sources and standards, according to detailed and specific rules

Quelle: [unintelligible]

8

Overall Process Map

Quelle: [unintelligible]

12

„minor defects“-Inspektionsraten

minor Insp.Raten

Quelle: [unintelligible]

14

Geschätzte Zusatzzeit für „Major defects“-S...

Zusatzzeit für Mj

Quelle: [unintelligible]

16

Selber ausprobieren ist noch überzeugender

Komplette Übungsspezifikation auf Testdokumenten
 + Länge: 1.63 p. + Länge: 2.2 p. + Systemstruktur: 1.53 p.

Quelle: [unintelligible]

17

Optische Täuschungen

Die Diagonalen sind parallel zueinander.
 Wenn es nicht genau, kann jemand nachrechnen. Bei diesem Bild darf man seiner Instanz nicht trauen!

Die optimale Inspektionsrate beträgt ca. 1 Seite pro Stunde.
 Wenn es nicht genau, kann jemand nachrechnen. Auch hier darf man seiner Instanz nicht trauen!

Quelle: [unintelligible]



19

Lesegeschwindigkeiten

Lesegeschw.

Quelle: [unintelligible]

23

Typische Inspektionsraten beim Review eines Testdokuments

typische Insp.Raten

Quelle: [unintelligible]

25

Wann findet ein Reviewer die Fehler in der Phase „Individual Checking“?

Quelle: [unintelligible]

27

Zeitpunkt der Fehlerentdeckung in Phase „Individual Checking“ - Reviewer

Quelle: [unintelligible]

28

Zeitpunkt der Fehlerentdeckung in Phase „Individual Checking“ - Inspektor

Quelle: [unintelligible]

29

Kumuliertes Histogramm für alle 28 Reviewer

Quelle: [unintelligible]

31

Defect Density against inspection time

Quelle: [unintelligible]

33

Ein Dokument mit 2 Methoden geprüft

Systemanforderliche mit 220 Seiten
 * geprüft mit herkömmlicher Review (alle 220 Seiten)
 * und geprüft mit Inspektion (nur 13-seitiges Kapitel)

Review	Methoden	Inspektionszeitpunkt	gefundene Fehler	insgesamt
herkömmliches Review	insgesamt: 23	insp. bei: 10 p.	4	max. 2%
Inspektion	insgesamt: 41	insp. bei: 4 p.	95	max. 23%

Quelle: [unintelligible]

36

Erfolgreiche Projekte sind nicht der Normfall

Quelle: [unintelligible]

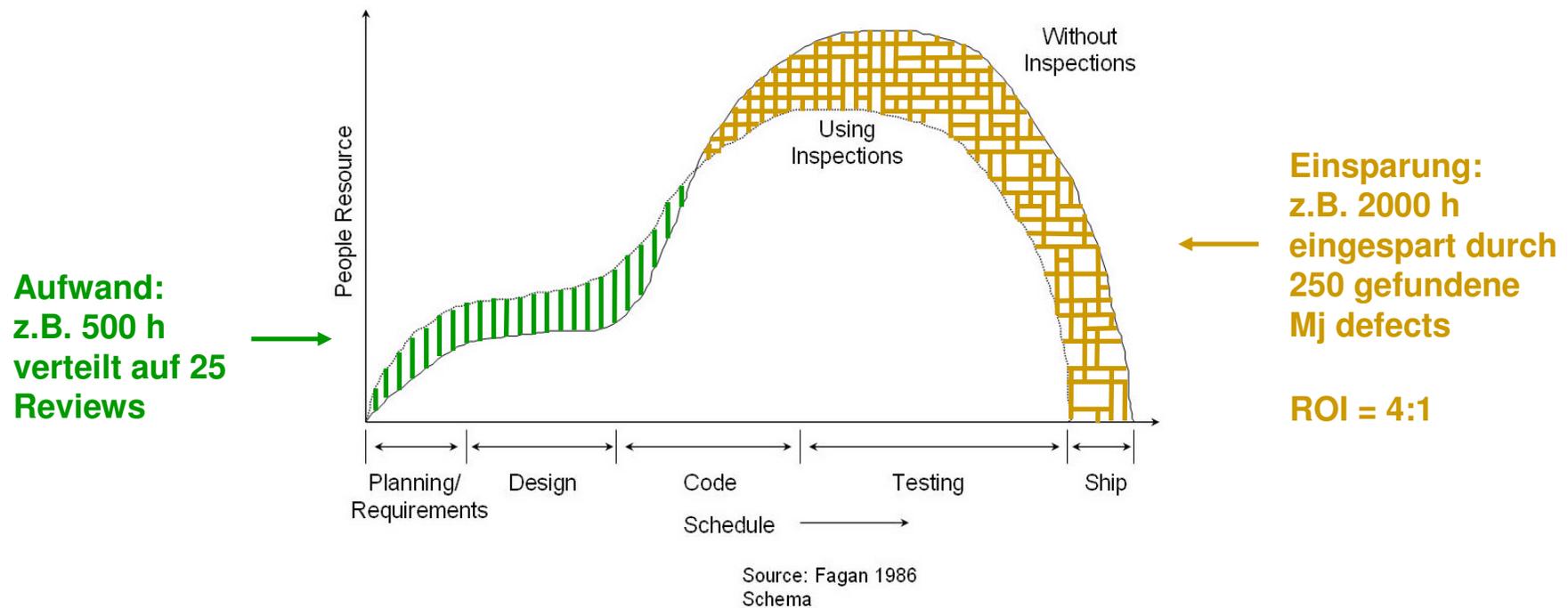
Es folgen Reserve-Folien für die Diskussion

39

Aufwand für 100%ige Reviewabdeckung

40

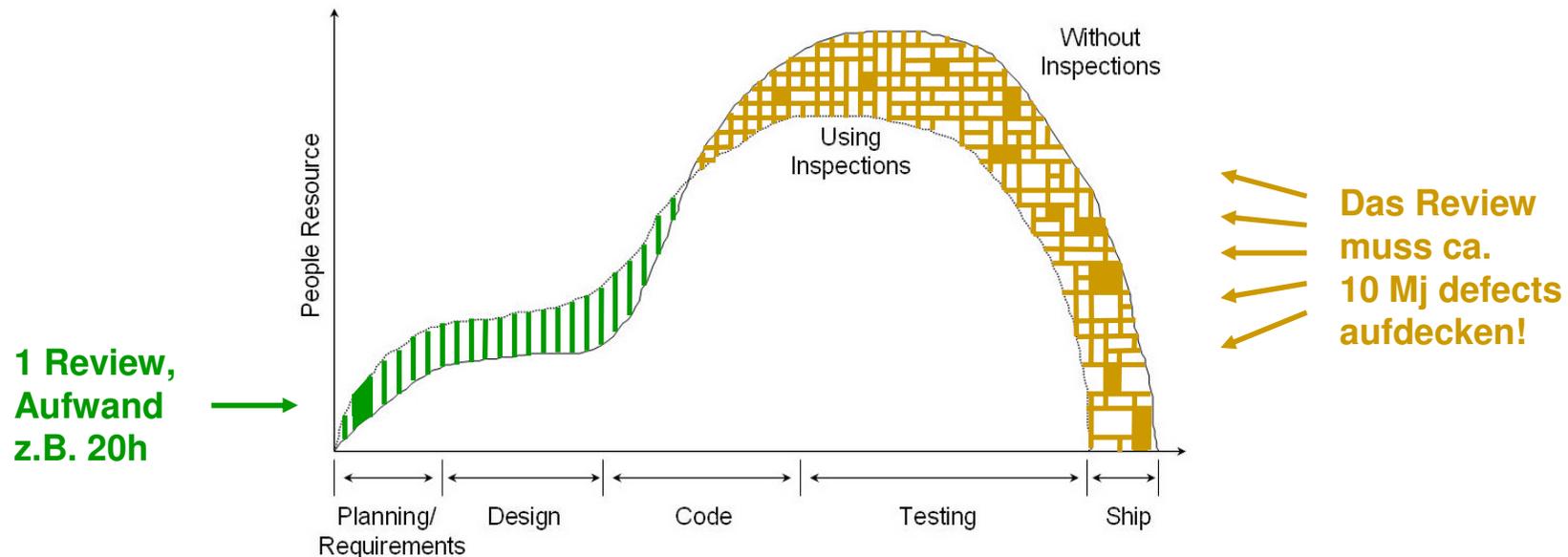
Gilb: ca. 10-15% des Projektaufwands (Radice: ca. 8-20%)



Keiner verlangt von Ihnen 10-15% Aufwand en bloc!

41

Sie investieren zuerst nur in 1 Review (Aufwand 10 - 20h)



Wurden genug Mj defects gefunden?

Ja \Rightarrow es folgt das nächste Review

Nein \Rightarrow nichts tun, die Qualität ist schon gut genug

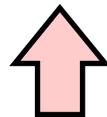


Beispiel zur optimalen Inspektionsrate

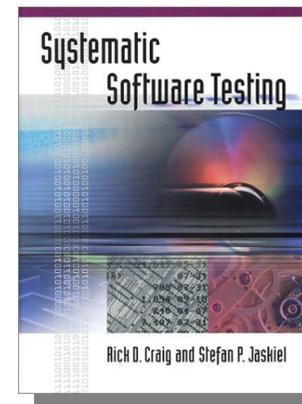
42

Automated Teller Machine (ATM) requirement:

»*A valid user must be able to withdraw up to \$200 or the maximum amount in the account.*«



**Diese 18 Wörter sind in ca. 4 - 5 Sek. gelesen
(entspricht ca. 50 p/h)
Zum Prüfen sind ca. 3 - 4 min. nötig
(entspricht ca. 1 p/h)**



Ref.: R. Craig, S.P. Jaskiel
Systematic Software Testing