



# **Effekte automatischer Bewertungen für Programmieraufgaben in Übungs- und Prüfungssituationen**

**Michael Striewe, Michael Goedicke**

**Universität Duisburg-Essen**

**Institut für Informatik und  
Wirtschaftsinformatik**

**<http://www.s3.uni-due.de/>**



## Agenda

- Automatische Prüfung von Programmieraufgaben
- Rahmenbedingungen der Untersuchung
- Qualität automatischer Bewertungen
- Allgemeine Effekte automatischer Bewertungen
- Effekte im Übungsbetrieb
- Effekte in Prüfungssituationen
- Fazit

# Automatische Prüfung von Programmieraufgaben

- Verschiedene Einsatzbereiche:
  - Intelligente Tutoren-Systeme
  - Rechnergestützte Prüfungen
- Verschiedene Techniken:
  - Statische Programmanalyse
  - Dynamische Programmanalyse
- Interessanter Untersuchungsgegenstand:
  - Offene Frageformen
  - Komplexer als Multiple-Choice
  - Auswirkung auf Lernende bisher kaum untersucht

## Rahmenbedingungen der Untersuchung

- Studierende in der Vorlesung "Programmierung" (1. Semester)
- Prüfungssystem JACK:
  - Serverbasiertes System vom Herunterladen von Aufgaben und Hochladen und Lösungen
  - Automatische statische und dynamische Prüfung jeder Lösung
  - Manuelle Nachkorrektur möglich
- Umfang:
  - 6 Testate unter Prüfungsbedingungen (je 45 Minuten)
  - 6 vorbereitende Übungsaufgaben
  - Insgesamt 2721 bewertete Lösungen
  - Umfrage mit 65 Teilnehmern

# Qualität automatischer Bewertungen

- Qualitätsmaße:
  - Wie oft muss das automatische Ergebnis manuell überschrieben werden?
  - Wie viele Fehler werden übersehen?
- Beobachtung:
  - 2% der automatischen Ergebnisse (53 von 2721) mussten manuell überschrieben werden
  - 94% der Befragten beobachteten "selten" oder "nie", dass inkorrekte Lösungen irrtümlich akzeptiert wurden
  - Aber: Lernende unterscheiden zwischen dem Auffinden und dem Erklären von Fehlern
- Beurteilung:
  - Qualitativ ausreichende automatische Bewertungen sind möglich

# Allgemeine Effekte automatischer Bewertungen

- Wichtige Fragestellungen:
  - Wird der (allgemein angenommene) Geschwindigkeitsvorteil tatsächlich als positiver Effekt gesehen?
  - Besteht eine grundsätzliche Skepsis gegenüber automatischen Bewertungen?
- Beobachtungen:
  - 75% der Befragten (49 von 65) sehen Geschwindigkeit "völlig" oder "weitgehend" als Vorteil und zeigen keine Skepsis gegenüber automatische Bewertungen
  - 86% der Befragten (56 von 65) halten das System allgemein für nützlich, wünschen sich aber weitere Verbesserungen
- Beurteilung:
  - Keine grundsätzlichen negativen Auswirkungen feststellbar

## Effekte im Übungsbetrieb

- Wichtige Fragestellungen:
  - Können alleine aus automatische Meldungen Lernfortschritte erzielt werden?
  - Ist die Arbeit mit automatischen Bewertungssystemen motivierend?
- Beobachtungen:
  - 88% der Befragten (57 von 65) brauchten "selten" oder "nie" zusätzliche Erklärungen zu den automatischen Meldungen
  - Das System motiviert und hilft bei der Lösung, aber nur bedingt beim Verständnis der Aufgaben
  - 77% der Befragten (50 von 65) betrachten das System im Übungsbetrieb als "völlig" oder "weitgehend" nützlich
- Beurteilung:
  - Viele positive Effekte, keine negativen Effekte

## Effekte in Prüfungssituationen

- Wichtige Fragestellungen:
  - Werden automatische Prüfungen als gerecht erachtet?
  - Erzeugen automatische Bewertungen zusätzlichen Stress?
- Beobachtungen:
  - 55% der Befragten (36 von 65) halten die automatischen Bewertungen für "völlig" oder "weitgehend" gerecht, allerdings nur 20% für gerechter als menschliche Prüfungen
  - Keine Tendenz bezüglich der Erzeugung von zusätzlichem Stress
  - 66% der Befragten (43 von 65) betrachten das System im Prüfungsbetrieb als "völlig" oder "weitgehend" nützlich
- Beurteilung:
  - Keine negativen Effekte, sondern Gleichwertigkeit zu manuellen Prüfungen



## Fazit

- Dasselbe System für Übungen und Prüfungen verwendbar
- Automatische Bewertungen mit manuellen Bewertungen vergleichbar
- Automatische Bewertung von Programmieraufgaben sinnvoll und gewinnbringend
- Erweiterung auf anderen Einsatzbereiche erstrebenswert