Automation Support for Feedback and Assessment in Learning Programming by Doing

For WMSEE 2025 on 2025-05-21

Arthur Rump
University of Twente

context

why and where do we want automated assessment?

assessment

assessment

formative

summative

assessment

formative   ## summative

"grades"

assessment

formative

"feedback"

summative

"grades"

assessment

formative

"feedback"

summative

"grades"

required

assessment

formative

"feedback"

actually useful

summative

"grades"

required

open-ended assignments

open-ended assignments

**TCS M2
Programming Project**

build a board game

client/server architecture

from scratch

open-ended assignments

**TCS M2
Programming Project**

build a board game

client/server architecture

from scratch

**CreaTe M4
Algorithms**

"build something"

interactive

use the building blocks

why is that hard?

don't build an autograder

use automation to *support* assessment

requirements

requirements

supporting

requirements

supporting

during assessment

requirements

supporting

during assessment

during configuration

requirements

supporting   **flexible**

during assessment

during configuration

requirements

supporting   **flexible**

during assessment   extensible

during configuration

requirements

supporting

flexible

during assessment

extensible

during configuration

configurable

how?

two perspectives on assessment

two perspectives on assessment

answering questions

two perspectives on assessment

answering
questions

gathering
evidence

gradual automation?

can we use manual assessment to configure automation?

combining manual and automated assessment?

automated → manual is easy?

what about *manual → automated*?

Encapsulation

sufficient
All fields except constants are private.
good, additionally:
Methods are only public when intended to be used from other classes.

Structure

Interfaces are applied where appropriate to ensure low coupling between components.

current work:

dsl for rubrics

to define interdependencies

not an autograder

not an autograder
a tool should be **supporting** and **flexible**

not an autograder
a tool should be **supporting** and **flexible**
through gradual automation?

not an autograder
a tool should be **supporting** and **flexible**
through gradual automation?
through a combination of manual and automated
assessment?

not an autograder
a tool should be **supporting** and **flexible**
through gradual automation?
through a combination of manual and automated
assessment?

how about LLMs?

requirements for assessment

reliability   validity   transparency

is it all bad?

no, but use it within a system