Evidence-Based Feedback in Higher Education through Constructive Alignment and Cognitive Diagnostic Modeling

Stefan Behrendt, B.Eng. M.Sc.

University of Stuttgart, Institute of Educational Science

June 8, 2021

Constructive Alignment provides a theoretical model for outcome-oriented teaching in higher education. Defining the learning objectives as skills and using the tasks from university exams, one can apply Cognitive Diagnostic Modeling methods. This contribution investigates practical implications of this approach for teaching and learning, focusing on feedback processes exemplary in mechanical engineering courses.

The only feedback systems implemented in higher education often remain point-based or letter grades, despite significant research about how to implement functional feedback recommending alternatives to grades-based scores (Carless, 2006). Despite there are even politically driven processes towards outcome-oriented teaching like in the European Union, teachers and students are not able to use them in a promoting way (Schaper et al., 2012).

This may be traced back on the lack of methods enabling teachers to gain deeper knowledge about their teaching and learning measurement procedures as well as to provide detailed feedback on the students' abilities.

1 Theoretical Background

1.1 Constructive Alignment

John Biggs introduced and developed a simple didactical model that is intended to align the three most important fields of didactical decisions in higher education: the *intentended learnings outcomes*, the *teaching and learning acitivties* and the *assessment tasks*. It is based on the assumptions of construcivism and therefore aims at outcome-oriented teaching (Biggs, 1996). Figure 1 shows the model called *Constructive Alignment* (CA).

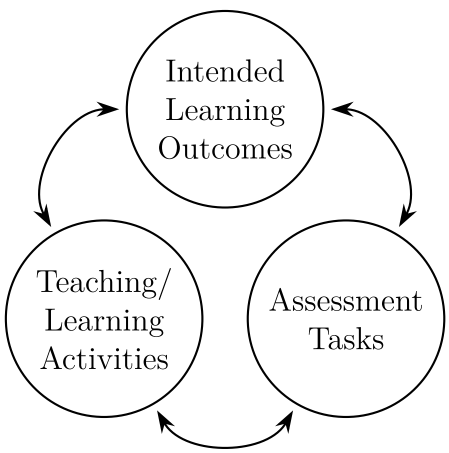

Figure 1: Model of *Constructive Alignment*

Biggs states that teachers often use learning aims instead of learning objectives, which do not correspond to their teaching activities. So, those aims often remain unattainable and are also not examined. He introduces a supporting model for helping teachers to break down their demands and to use learning objectives which can be achieved by the students: the *SOLO taxonomy* (Biggs, 1996). Applying this taxonomy, learning outcomes can be aligned with the teaching and learning activities and can be examined by the tasks in an assessment. Figure 2 shows the five levels of understanding he deduced as well as their hierarchical structure.

Since the defined objectives are transparent to the students, they are able to check by themselves whether the teaching and learning activities fit to the teacher's demands (Biggs, 1996).

1.2 Competence Orientation

Even though Biggs does not use the term *competence*, his model is very close to this concept through the usage of *constructivism* and *outcome*-orientation. There are numerous papers within the field of higher education didactics that use the model of CA as a basis for implementing competence-oriented teaching.

Nevertheless, Schaper et al. (2012) stated that an explicitly stated competence model in higher education is often missing. They want to ease the process of relying on such a model by defining some basic features of academic competences (Schaper et al., 2012):

- They are reflexive and explicable as well as evidence-based.
- They are organized by scientific domains.
- They refer to complex and new situations and tasks.
- They are set in a specific field of activities.

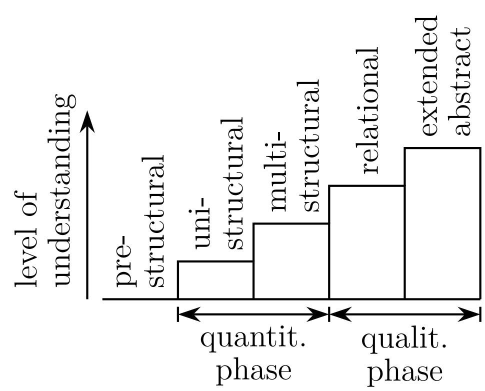

Figure 2: Model of *SOLO taxonomy*

1.3 Feedback in Higher Education

In most cases, the only feedback students receive in higher education exams are grades and in some instances individual inspections. It has been shown that "students are often dissatified" (Carless, 2006) with this type of feedback. The three main reasons identified are:

- it is "difficult to interpret"
- it is "lacking specific advice to improve"
- it has "a potentially negative impact on students' self-perception and -confidence"

(Carless, 2006)

Hattie and Timperley (2007) develop a simple feedback model that is able to address the stated criticism by explicitly regarding three steps as shown in figure 3.

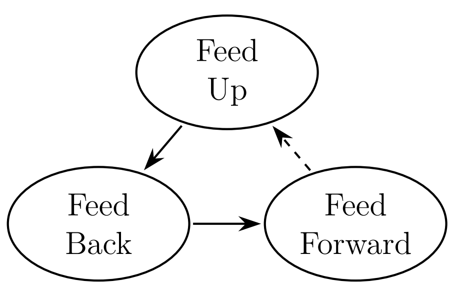

Figure 3: Model of *Feedback*

They extend the feedback in the narrow sense by an upstream process of defining the individual goals and a downstream process of interpreting the feedback for individual improvement (Hattie & Timperley, 2007).

1.4 Cognitive Diagnostic Models

Cognitive Diagnostic Models (CDM) "promote assessment for learning rather than assessment of learning" (Ravand & Robitzsch, 2015). It belongs to the class of *Linear Logistic Test Models* and thus is a probabilistic approach.

Some of the simplest models within this approach are the *(non)compensatory deterministic input noisy-and-gate models* (DINA resp. DINO). They bring together a skill space comprising dichotomous skills ($a \in \{0,\ 1\}^D$) with dichotomously coded tasks ($x_i \in \{0,\ 1\}$). A *q-matrix* defines which skills need to be mastered to solve a task correctly. DINA defines that all assigned tasks need to be mastered, DINO respectively at least one of the assigned tasks. Both assumptions can be mixed for modeling a test (von Davier & Lee, 2019).

For each task (i) a guessing (g_i) and a slipping (s_i) parameter are estimated and for each person (j) the individual skill profile (a_j), using the equation

$$\begin{aligned} P(x_i|a_j) &= \left((1-s_i)^{\xi_{a_j q_i}} g_i^{1-\xi_{a_j q_i}}\right)^{x_i} \\ &\cdot \left(s_i^{\xi_{a_j q_i}} (1-g_i)^{1-\xi_{a_j q_i}}\right)^{1-x_i} \end{aligned}$$

which is mathematically the simplest way to bring the parameters together (von Davier & Lee, 2019).

2 Research Model

2.1 Basic Model

Biggs (2003, p. 3) demands that "[w]e have to state our objectives in terms of what we want students to do". Regarding the definition of *mastery models* (von Davier & Lee, 2019), like DINA or DINO, we can translate this into suitable mathematical terms: $a_{v,k} = 1$ if person v masters learning objective k and $a_{v,k} = 0$ if not. Therefore, the *skill space* is defined by the learning objectives of *CA*.

Furthermore, Biggs (2003, p. 5) demands that "[t]he score an individual obtains reflects how well he meets preset criteria". This as well perfectly meets the definition of *mastery models* as shown in table 1 regarding the additional assumption that $1 - s_i \geq g_i$ to have the scale directed correctly.

Table 1: The four states in a *mastery model* (von Davier & Lee, 2019, p. 7)

$P(\rightarrow \mid \downarrow)$	$X_i = 0$	$X_i = 1$
$a = 0$	non-master	guessing
$a = 1$	slipping	master

Therefore, there is a mathematical model that represents the link between learning objectives and exam tasks within the theoretical model of *CA* sufficiently.

2.2 Research Question

Applying Hattie's feedback model and assuming that teachers want to follow the *CA*-approach, they ask themselves: Do the exam tasks match the learning objectives? What has to be changed to improve the match? Have the changes been successful regarding this aim? This is collected within the first research question:

> *Q1:* Can *Cognitive Diagnostic Modeling* contribute to apply evidence-based instructional design in the context of *Constructive Alignment*?

Additionally, Hattie's feedback model can be applied to the students' view: The *feed up* is to make the learning objectives understandable and transparent as it is stated within *CA*. Then the *feed back* has to deliver information about the individual ability to master those learning objectives. The *feed forward* has to provide information how to improve the ability. This is addressed within the second research question:

> *Q2:* To what extend helps *Cognitive Diagnostic Modeling* to provide the feedback to students for further learning as required in *Constructive Alignment*?

3 Methods

3.1 Model and Fit

As will be described in section 3.2, each task refers to one single learning objective. So, it is not important to differentiate between the *DINA*- and the *DINO*-approach (Junker & Sijtsma, 2001). The justification of the usage of *mastery models* to fit the purposes of modeling the theory of *CA* is described in section 2.1.

Since those models are special cases of *Linear Logistic Test Models*, the tasks are not allowed to be locally dependent (von Davier & Lee, 2019). This is checked using a statistic based on the adjusted residual correlation: $Q_{3,*}$ (Chen & Thissen, 1997). In order to preserve most items, the cut-off value is oriented at 0.5.

The additional assumption that $1 - s_i \geq g_i$ is implemented manually by deleting all items not fulfilling it.

As an absolute fit measure the *Standardized Root Mean Square Residual* (SRMSR) is regarded (Maydeu-Olivares, 2013). It describes the proportion of the data that is not explained by the model and therefore the misfit in applying *CA*.

All analyses are performed using R (R Core Team, 2021) and the package CDM (Robitzsch et al., 2020). This paper additionally uses the package knitR (Xie, 2021).

3.2 Sample

This paper uses data from a university lecture called *Design Theory I+II* within the course of study *Mechanical Engineering* at the University of Stuttgart, Germany. Its classes are located during the first and second semester and the exam is written after the first year of study.

The **Exam** is divided into three parts which are using different measurement approaches:

pt. A: Comprehensional Knowledge 20 min

- 40 Single-Choice-Items w/4 options

pt. B: Designing and Drafting 50 min

- 1 draft using 44 evaluation criteria

pt. C: Design Calculation 50 min

- 36 factors within 6 major contexts

This structure captures the interpretation of *competence* used by vocational eduation researchers which distinguishes *knowledge*—including its comprehension—and its *application*—in this subject operationalized by drafting and calculating—and therefore follows the European Union's demands on compentence orientation in higher education (Schaper et al., 2012).

Each item, evaluation criterion and factor is coded dichotomously. The calculation contexts as well as the different measuring approaches are not relevant for the statistical modeling (Behrendt, in preparation).

The **Learning Objectives** were created by the responsible institutes—*Institute for Engineering Design and Industrial Design* and *Institute of Machine Components*—refering to the didactical expertise of the *Center for Higher Education and Lifelong Learning*. They are available publicly within the module's description[1].

These learning objectives are the basic of the teaching and learning acitivities and follow the *SOLO taxonomy*.

Each item, evaluation criterion and factor adresses precisely one learning objective, since those objectives are encompassing to ensure a manageable set within the broad topic.

The assignment of each item to each learning objective has been carried out by an expert of the *Institute of Educational Science* under control of the responsible institutes. Table 2 shows the final Q-matrix in an aggregated way using the number of items of each part that correspond to each learning objective. Learning objectives having only 3 or less representatives are combined to similar ones.

The **Sample** consists of all students that took the exam ($N = 333$) in the year 2016. 14 % are females, which is not uncommon for STEM courses in Germany (Autorengruppe Bildungsberichterstattung, 2020, p. 191). All items are measuring equal regarding the sex (Behrendt, in preparation).

[1]current description available under C@MPUS

Table 2: Aggregation of the Final Q-Matrix

LO	pt. A	pt. B	pt. C	$\sum$
1	0	7	0	7
2	0	0	0	0
3	4	12	0	16
4	3	0	0	3
5	2	0	0	2
6	3	15	0	18
7	2	0	6	8
8	10	1	18	29
9	2	0	0	2
10	3	9	0	12
11	11	0	12	23
12	0	0	0	0

There are two main courses of study: *General Mechanical Engineering* ($n = 203$) and *Vehicle and Engine Technology* ($n = 126$), a specialization within mechanical engineering. All items also are measuring equal regarding these two groups (Behrendt, in preparation).

In order to validate the interpretation of the results, they are compared to the final grades of the exam. Figure 4 shows their distribution.

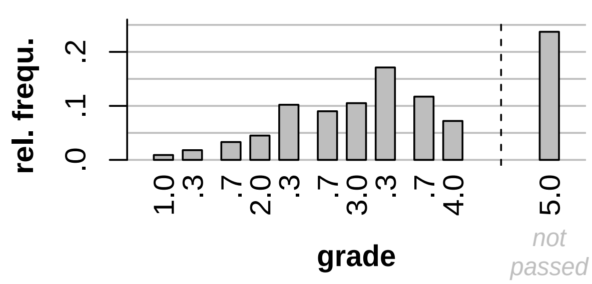

Figure 4: Distribution of the Grades

4 Results

4.1 Model Fit

4 items were deleted due to $1 - s_i < g_i$. Another 6 items were deleted due to local item dependency. So 109 items (92 %) remain.

With an SRMSR = .086 the absolute model fit is nearly acceptable. The remaining items do not show any relevant local item dependency ($Q_{3,*} \in [-0.31,\ 0.54]$), having a small variation (MAD $Q_{3,*} = 0.053$).

4.2 Estimated Parameters

Figure 5 shows a bar with a minimum g_i and a maximum $1 - s_i$ for each item i, grouped by the learning objectives. Most items show a sufficient discrimination, but there are some conspicuities that needed to be discussed with the subject experts.

The mastery rates of each learning objective shown in table 3 identify a comprehensible variation in the difficulty of mastering them. The minimal mastery rate is .39, the maximal mastery rate is .65.

Table 3: Mastery Rates

1	3	6	7	8	10	11
.65	.41	.50	.51	.65	.39	.50

The correlation structure shown in table 4 indentifies only moderate relations between the learning objectives, even between those with close mastery rates. The minimal correlation is .10, the maximal correlation is .51.

Table 4: Correlations

	3	6	7	8	10	11
1	.29	.31	.24	.16	.22	.29
3		.32	.23	.22	.40	.38
6			.36	.26	.51	.32
7				.11	.21	.10
8					.19	.33
10						.31

4.3 Validation on Grades

Figure 6 shows the relations between mastering a learning objective and the achieved grade. The diamonds visualize the means of the students having mastered respectively not mastered the learning objective. For those having not mastered the learning objective the highest grade achieved is marked by a line. Additionally, the point-biserial correlation is displayed.

Each single learning objective shows high discrimination regarding the grades. The means differ by at least one grade level. There are few people that master some learning objectives but do not pass the exam. Furthermore, there is no student with best grade who does not master all learning objectives.

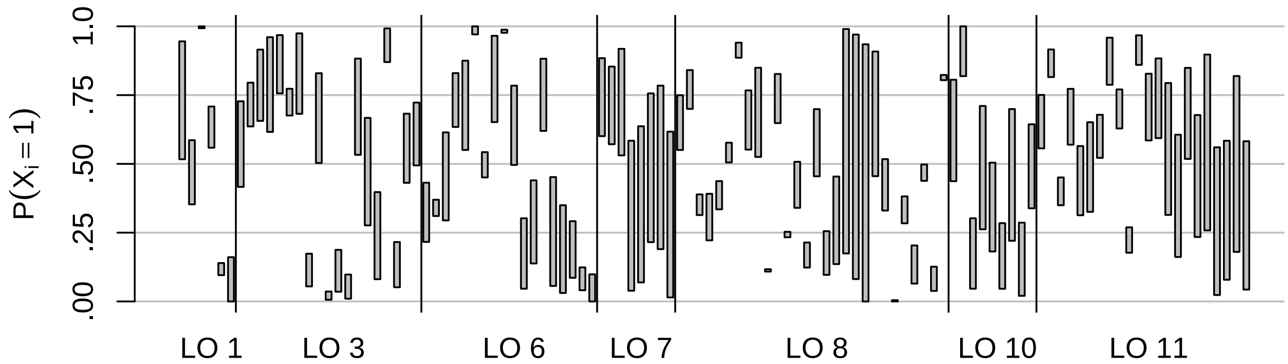

Figure 5: Slipping and Guessing Parameters grouped by Learning Objectives

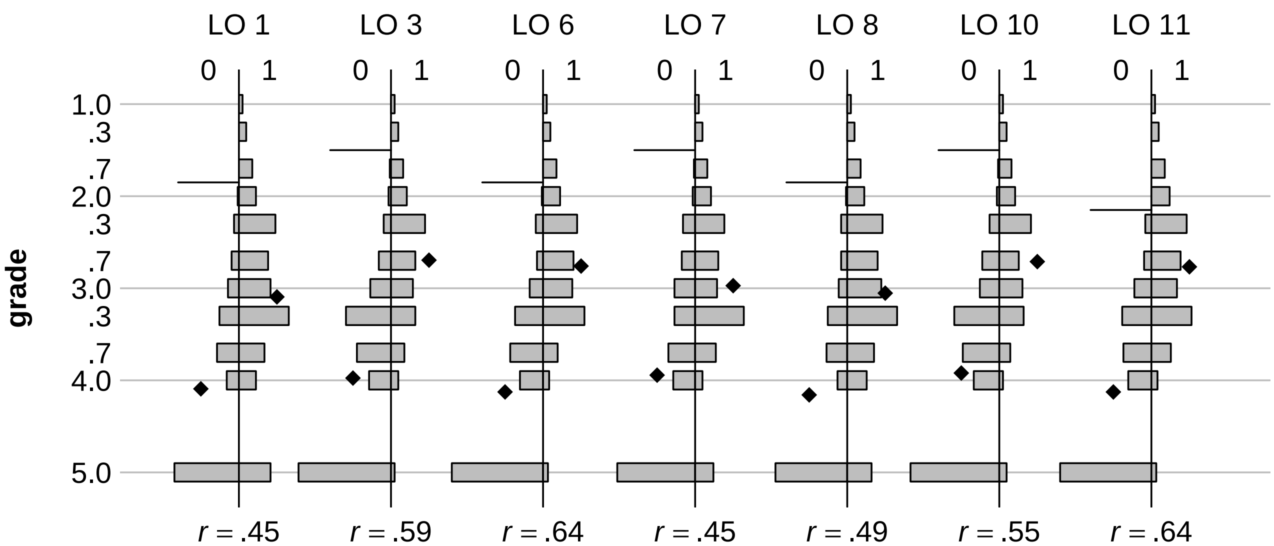

Figure 6: Relations between Mastering a Learning Objective and Grades

5 Discussion

5.1 Question 1

Cognitive Diagnostic Models show a parametrization which is easy to understand and to interpret, even for subject experts who are not familiar with psychometric methodology. Since there are not many assumptions that need to be taken into account, most items could be kept and can be inspected in detail.

Regarding the difference between guessing and slipping parameters, some items can be identified that do not fit to the others. E.g. there is one evaluation criterion in the draft that captures the care in drafting. This is an important criterion but does not represent abilities that fit to the other criteria or that are covered by the learning objectives.

The item pairs that are concerned by *local item dependency* were not surprising. E.g. there is an evaluation criterion that a special part of the draft has been regarded and some other criteria that evaluate the quality of that part.

Therefore, it is not necessary to eliminate those items in future exams. But the subject experts produced some further interesting findings they were not aware of. Regarding the dependencies they identified e.g. that the requirements of technical mechanics can not be included in the calculation contexts even if they master those requirements.

Furthermore, building the Q-matrix showed the experts the different weighting of the learning objects within the exam. Learning objective 2 (statial thinking) and 12 (using CAD software) can not be measured using a written exam. But e.g. learning objective 4 (knowing and using standards) is underrepresented regarding its importance.

This way of modeling also showed that not the calculation—learning objectives 8 and 11—but the drafting—learning objectives 3 and 10—is the most difficult task for the students. So the teaching and learning activities can be optimized even more targeted.

5.2 Question 2

The validation on the grades show that the results of mastering the learning objectives can be interpreted comparably to the grades. This is important to ensure that there do not arise doubts about the quality of the grading system and to prevent misinterpretations of the additional feedback.

Since all students that reached the grades 1.0 and 1.3 have mastered all learning objectives, this model does not offer additional information to the highest achieving students. But for all other students it is much more informative with regard to the learning objectives which have to be worked on in more detail than only receiving the grade.

5.3 Perspectives

The learning objectives are formulated to the step of the *SOLO taxonomy* which should be reached, but some items measure the steps below. Therefore, it would be better to include this hierarchy in the model. This is a challenging task and has not been finished until now, but it may clearly improve the quality of the interpretation.

This paper presented the sample of only one year. There is data available for a sequence of six years of the course *Design Theory I+II* and a sequence of four years of the course *Design Theory III+IV*. This can be used to validate possible interpretations of the absolute fit index as well as to validate if the interpretations are stable across the years.

Linking the results of different courses offers the possibility of estimating the forecasting power of mastering the learning objectives which can help to motivate the students to learn unmastered topics.

5.4 Further Potentials

Since the modeling process is relatively easy, there is a high potential for automization. This could be used to not only model exams but also all kinds of excercises and working sheets. So, unmastered learning objectives can be identified earlier and the students can improve their abilities before not passing the exam. This could prevent for study dropout, which is an important problem.

Additionally, the monitoring of different fit measures can be used to track educational quality, regarding the teaching and learning activities as well as the quality of the exam itself. Latter is very important when changing the exam mode, e.g. to online assessment.

References

Autorengruppe Bildungsberichterstattung (Ed.). (2020). *Bildung in Deutschland 2020: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. wbv Publikation.

Behrendt, S. (in preparation). *Erweiterung der Aussagekraft der Leistungsrückmeldung durch Klausuren im Studium mittels Constructive Alignment und Cognitive Diagnostic Modeling* (Doctoral dissertation). University of Stuttgart.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*, 347–364.

Biggs, J. (2003). *Aligning Teaching and Assessment to Curriculum Objectives* (tech. rep.). Imaginative Curriculum Project, LTSN Generic Centre.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, *31*(2), 219–233.

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265–289.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models (with discussion). *Measurement: Interdisciplinary Research and Perspectives*, *11*, 71–137.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ravand, H., & Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. *Practical Assessment, Research and Evaluation*, *20*, 1–12.

Robitzsch, A., Kiefer, T., George, A. C., & Ünlü, A. (2020). *CDM: Cognitive Diagnosis Modeling* [R package version 7.5-15]. https://CRAN.R-project.org/package=CDM

Schaper, N., Reis, O., Wildt, J., Horvath, E., & Bender, E. *Fachgutachten zur Kompetenzorientierung in Studium und Lehre*. 2012. https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf

von Davier, M., & Lee, Y.-S. (2019). *Handbook of Diagnostic Classiffcation Models*. Springer International Publishing.

Xie, Y. (2021). *knitr: A General-Purpose Package for Dynamic Report Generation in R* [R package version 1.33]. https://yihui.org/knitr/

stefan.behrendt@ife.uni-stuttgart.de