Intersectionality and Incremental Value: What Combination(s) of Student Attributes Lead to the Most Effective Adaptations of the Learning Environment?

Jenna Olsen [0000-0002-4545-4976] and Sydney Shackelford [0000-0002-9860-7729]

Western Governors University, Salt Lake City UT 84107, USA
lncs@springer.com

Abstract. Students can be placed in more than one category at the start, middle, and end of their educational journey. These categories can be based on demographics (age, gender, sex, minority, disability, ethnicity), on behavior (procrastination, struggle, frustrated guessing, pathological re-reading), on individual attributes (help-seeking, locus of control, time management, optimism), on community (internet access, setting, average education and income), and on academic factors (previous grades and degrees). These categories are frequently used by faculty, designers, and leadership to seek a better understanding of students and their needs with the goal to personalize or adapt the learning environment in the hopes of leading to more effective learning and more successful student outcomes. In these analyses we seek to determine the relative value of different student categories – and how these can be combined to result in the most effective educational process. We find ourselves asking what attributes matter the most – and which interact with each other to increase or reduce the amount of relative value. It is worth noting that several of the categories – while they play a large role in our students' holistic selves – are both highly sensitive (frequently protected) and static. If we can approach or match their value (educationally) in other categories which are less sensitive or more changeable, that will be a positive result. Because, while these attributes play a role in who the students *are*, they need not play a role in how the students are *taught*.

Keywords: Student Attributes, Adaptive Learning, Behavior.

1 Introduction

1.1 Potential Student Categories

Students can be placed in more than one category at the beginning – and throughout – their educational journey. These categories can be based on demographics (age, gender, sex, minority, disability, ethnicity), on behavior (procrastination, struggle, frustrated guessing, pathological re-reading), on individual attributes (help-seeking, locus of control, time management, optimism), on community (internet access, setting, average education and income), and on academic factors (previous grades and degrees).

Students exist in all these categories simultaneously, but the categories are not of equal value to educational adjustments.

These categories are frequently used by faculty, designers, and leadership to seek a better understanding of students and their needs – to personalize or adapt the learning environment in the hopes of leading to more effective learning, and more successful student outcomes.

In these analyses we seek to determine the relative value of different student categories – and how these can be combined to result in the most effective educational process. With this in mind, we find ourselves asking what attributes matter the most – and which interact with each other to increase or reduce the amount of relative value.

It is worth noting that several of the categories – while they play a large role in our students' holistic selves – are both protected and highly sensitive. If we can approach or match their value (educationally) in other categories which are less sensitive, that will be a positive result. It is also worth noting that many of the categories we evaluate are not changeable (or at least extremely difficult to change) including things like community, sex, minority, and disability. Again, these play a large role in who are students are, but they need not play a large role in how we educate them.

2 Courses and Student Attributes

2.1 Course Selection

Our analysis looks at two different courses – both offered at Western Governors University to different individuals in the student body. There is little, if any, overlap in the student population for the two courses. The data for both courses spans the same time frame – from 2017 through 2020. The criteria for course selection included resource data availability, assessment design, and student size.

Both courses use the same set of learning resources – so the data for student activity were well-matched and accessible. Both courses use an assessment style referred to as an objective assessment. This style uses individual question items where aspects such as validity, reliability, and security are carefully maintained by the psychometric team. This makes it easier to track test outcomes at the item, objective, and competency levels.

Finally, both courses have over 25,000 students in the data set, which allows us to test out various attributes, and attribute groups, for our clustering while maintaining large enough n-counts in the final results to trust our initial findings.

The first of these courses is C785 – Biochemistry (n = 45,445) which is used by students in the College of Health Professions. These students have some amount of nursing training and are attending the university to increase their credentials in order to further their career. This is an extremely difficult course within the program – it is not unheard of for students to retake the course once or twice in order to pass.

The second course is C165 – Integrated Physical Science (n = 26,816). This is a general education course used by students from the other colleges (Health Professions students are – very much – the exception in this student population). This course focuses on chemistry, physics, and earth science.

WGU Course and Term Structure. Students at Western Governors University register for academic terms of a six-month duration. A term starts every month (rolling terms) and students may be in two terms over the course of the year. Within these terms, they can be registered for a wide range of courses – counts from 4 to 10 are not uncommon – which they are able to complete at their own pace. It is expected that students will focus on one course at a time during their term, but this is not strongly enforced.

2.2 Student Attributes

For this study we focused on three broad categories of attributes related to the students in the courses.

Student Activity or Behavior. These attributes include the number of days logged in, number of courses or resources accessed, and assessment attempts. Throughout this study these attributes are measured at various points in time, including 7, 14, and 21 days into a six-month academic term. All measures for these attributes are cumulative from the start of the term through the day of measurement.

Student Readiness. These attributes are tied to the individual students' learning styles, technical and reading skills, and individual attributes. The assessment for these attributes has been used by the Health Professions college throughout the duration of our study.

Student Demographics. This category, which we frequently called 'Day 0' variables, includes demographic and personal information we collect from students before they start their first term. This includes self-reported ethnicity, pell awards for the first year of study, gender, and student zip code – which we use to connect to census information for the community and state of student residence.

Student Outcomes. The student outcome variable of interest is whether they passed the course during that enrollment, which we define as the combination of a student, a term, and a course. This is a binary outcome as students will either pass of rail the course during that enrollment. It is important to note that our analysis specifically excluded those students who were able to effectively 'test out' by passing the final assessment without engaging either with the course materials or with faculty.

3 Attribute Selection – C785 Biochemistry

3.1 Evaluation of Attributes

One dataset each for activity (measured at day 21), readiness, and demographics was loaded into WEKA explorer. This was done in order to search for the best attributes – those with the highest differentiation in student outcomes - in each dataset which would then be used to create student clusters. Multiple attribute selection methods were used for each dataset to identify commonalities and select attributes that were highly ranked across multiple methods.

For each dataset, two select attribute methods were used. These were BestFirst search with the CFSSubsetEval evaluator and the Ranker search with CorrelationAttributeEval evaluator.

C785 Activity Attributes. The selected attributes within the activity category were *days_engaged* and *links_per_day*. The attribute *days_engaged* is defined as the number of days during which the student engaged in *any* course during the first three weeks of the six-month term (through day 21). Links per day is the number of links – again in *any* course – that the student clicked. These links guide students to both required and supplementary resources and to hosted course content.

Table 1. Activity Attribute Clusters – C785

Cluster	Days Engaged	Links Per Day	Description
Cluster 0	15	2.2	Highly engaged students who check in with their coursework often
Cluster 1	10.3	1.8	Engaged students who may use some, but perhaps not all, of the links.
Cluster 2	7.0	1.6	Average engagement and use of course links
Cluster 3	1.9	0.9	Very low engagement during the early weeks of the term
Cluster 4	4.7	1.0	Moderate engagement with low use of links

C785 Readiness Attributes. Three attributes were selected by our methods from the readiness category. These were *personalattpct*, *readingpct*, and *procrastpct*. The *readingpct* attribute is a measure of the students' relative reading accuracy. The *personalattpct* is an average of all of the personal attributes measured by the survey. Finally, the *procrastpct* is a measure of procrastination tendencies. For all three of these attributes, a higher score is more desirable.

Table 2. Readiness Attribute Clusters – C785

Cluster	*personalattpct*	*readingpct*	*procrastpct*	Description
Cluster 0	88.0%	79.2%	95.3%	A student with above average reading accuracy who scores high in all personal attributes but especially well in proactive behavior.
Cluster 1	80.2%	61.6%	76.3%	A student with low reading accuracy and room for improvement in proactive behavior while average overall in all personal attributes.
Cluster 2	78.3%	87.2%	69.7%	A student who procrastinates but has high reading accuracy and room for improvement all around in personal attributes.
Cluster 3	83.8%	87.3%	81.7%	Student who has high reading accuracy and scores quite well in all personal attributes including proactive behavior.
Cluster 4	71.9%	78.9%	55.6%	A significant procrastinator who scores well in reading accuracy and has room for improvement in all personal attributes.

C785 Demographics Attributes. Three attributes were selected from the demographics category. These were *marital_status*, *household_income*, and *minority*. The *marital_status* attribute is a self-reported variable and can be one of six possible categories – Married, Divorced, Separated, Widowed, Single, No Response. The *household_income* is also a range of categories, starting with '< $16,000' and ending with '> $65,000'. Finally, the *minority* is a binary status of "yes" or "no".

Table 3. Demographics Attribute Clusters – C785

Cluster	*marital_status*	*household_income*	*Minority*	Description
Cluster 0	Married	Not reported	N	Married, non-minority students with unknown income levels
Cluster 1	Married	$45000-$64999	N	Married, non-minority students with above-average incomes
Cluster 2	Single	$65000+	N	Single, non-minority students with the highest reported incomes
Cluster 3	Married	$65000+	N	Married, non-minority students with the highest reported incomes
Cluster 4	Single	$45000-$64999	Y	Single, minority students with above-average incomes

4 Cluster Creation and Performance – C785 Biochemistry

4.1 Clustering Method – C785

The next step was to cluster the students based on the selected attributes in each of the four categories. Each dataset was uploaded to WEKA Explorer and clustered using KMeans method. For the purposes of this study, we used five clusters for each dataset.

For each cluster group the following were recorded; number of clusters, minimum and maximum percent of total student population, spread in percent of total population, minimum and maximum course completion rate, and spread in course completion rate.

The binary outcome variable of course completion was averaged into course completion rate across the cluster for the latter two calculations. A greater spread in the course completion rate percentage indicates greater cluster differentiation and goodness of fit.

4.2 Clustering Results – Single Attribute Categories C785

Table 4. Individual Attribute Category Performance – C785

Category	Min % of Population	Max % of Population	Spread in Population	Min % Completion	Max % Completion	Spread in Completion
Activity	8.5%	38.4%	29.9%	56.7%	92.5%	35.5%
Demographics	9.5%	36.4%	26.9%	66.9%	80.4%	13.5%
Readiness	14.2%	23.8%	9.5%	60.7%	76.1%	15.4%

For individually clustered categories of attributes, the activity category shows the greatest differentiation in both population spread, and outcome spread. The demographics category has the next largest spread in population, while the readiness category has the next largest spread in outcomes.

4.3 Clustering Results – Two Attribute Categories C785

To evaluate the attribute categories as combinations of two, rather than individually, we cross-matched students to two of their three cluster assignments. Then we repeated the queries from the individual category performance to measure population and outcome spread.

Table 5. Mixed Attribute Category Performance (two of the three categories) – C785

Category	Min % of Population	Max % of Population	Spread in Population	Min % Completion	Max % Completion	Spread in Completion
Activity & Readiness	1.1%	9.1%	7.9%	57.7%	95.5%	37.8%
Activity & Demographics	1.0%	12.1%	11.2%	52.2%	94.0%	41.8%
Readiness & Demographics	1.4%	9.3%	7.9%	64.0%	86.5%	22.5%

By cross-matching the students, we find several pieces of information.

The first is that the combination of activity and any other category of attributes results in a stronger cluster differentiation. It is worth noting that this spread represents a very significant growth in outcome differentiation for the other category (either demographics or readiness) with a much smaller growth for activity outcome performance. Activity improves the outcome differentiation of both the demographic (28.3% gain) and readiness (22.6% gain) categories.

The second is that we see that the combination of readiness and demographics clustering is stronger in outcome differentiation than either of the two categories separately, though still not a greater fit than activity either individually or combined with other categories.

4.4 Clustering Results – Three Attribute Categories C785

The final step in our evaluation of the different attribute categories for the biochemistry course was to cross-match students to all three of their clusters. By doing so, we are effectively clustering based on all three broad categories at once. We then ran, one additional time, the same analyses to measure both population and outcome differentiation.

Table 6. Mixed Attribute Category Performance (all categories) – C785

Category	Min % of Population	Max % of Population	Spread in Population	Min % Completion	Max % Completion	Spread in Completion
Activity, Readiness & Demographics	0.1%	3.4%	3.2%	40.0%	100.0%	60.0%

From these results we see that the combination of all the categories results in the strongest outcome differentiation at a 60% spread. We also see that the individual clusters represent a significantly smaller percentage of the total student population, with the smallest of the new clusters representing a group of less than 50 students, even though we started with a large population.

Adjustment to Cluster Planning based on C785 Results. Our analysis of the biochemistry course provided results which were unexpected, but very intriguing. After evaluating the first of the two courses, we adjusted our planned analysis for the integrated physical science course. Having found activity to be the strongest – by far – of the individual categories of attributes for the biochemistry course, we created additional categories of the activity attributes for the physical science course. We also added an additional question at this point in the analysis – specifically when (measured as number of weeks into the six-month term) did the value of activity data surpass that of demographics in differentiation of both student population and successful outcomes.

5 Attribute Selection – C65 – Integrated Physical Science

5.1 Evaluation of Attributes

Three separate datasets – all for activity – were built for the C165 course data. Datasets were built for student activity at 7, 14, and 21 days (or 1, 2, and 3 weeks respectively). Data cleansing process again included the removal of data for students who “tested out” of the course. Additionally, for students who took the course multiple times, we focused only on the first enrollment. Each data set was then uploaded to WEKA in order to select the most appropriate attributes. The attribute selection methods were the same across each dataset.

A dataset was also constructed for demographic data. Census data for community information was matched with student's personal data via their zip code. This data was then uploaded to WEKA to select attributes.

Due to the fact that CHP students were not likely to take the physical science course, and the historical readiness assessment data was for CHP students only, we did not create a cluster for readiness.

C165 Activity Attributes. For each dataset, two select attribute methods were used. These were BestFirst search with the CFSSubsetEval evaluator and the Ranker search with CorrelationAttributeEval evaluator. For C165, as for the earlier clusters with C785, the selected attributes within each of the activity categories were *days_engaged* and *links_per_day*.

Table 7. Activity Attribute Clusters – Week 1 – C165

Cluster	Days Engaged	Links Per Day	Description
Cluster 0	1.6	0.3	Very low engagement and use of links
Cluster 1	3.5	0.5	Low engagement
Cluster 2	7.5	1.2	High engagement, average links
Cluster 3	6	0.9	High engagement, average links
Cluster 4	4.9	0.7	Average engagement and links

Table 8. Activity Attribute Clusters – Week 2 – C165

Cluster	Days Engaged	Links Per Day	Description
Cluster 0	13.3	1.2	Very high engagement and use of links
Cluster 1	10.0	0.8	High engagement, average links
Cluster 2	2.9	0.3	Very low engagement and links
Cluster 3	7.6	0.7	Average engagement and links
Cluster 4	5.5	0.5	Low engagement and links

Table 9. Activity Attribute Clusters – Week 3 – C165

Cluster	Days Engaged	Links Per Day	Description
Cluster 0	8.5	0.5	Low engagement and links
Cluster 1	20.1	1.3	Very high engagement and links
Cluster 2	16.5	1.0	High engagement and links
Cluster 3	12.3	0.8	Average engagement and links
Cluster 4	4.4	0.3	Very low engagement and links

C165 Demographic Attributes. Two feature selections were run to find the best fit attributes. These were both Ranker search methods with the CorrelationAttributeEval evaluator and the ReliefAttributeEval evaluator. The selected attributes within the demographics category were *pct_wo_internet*, *ethnicity*, and *nonresident_alien*. The *pct_wo_internet* attribute is a community-level descriptor of the percent of those living without internet. It is extracted from census data using the student zip code. Both the *ethnicity* and *nonresident_alien* are self-selected by the student during their initial enrollment process. Ethnicity is categorical - and students are able to select multiple categories as desired. Nonresident alien is a binary condition.

Table 10. Demographic Attribute Clusters – C165

Cluster	*pct_wo_internet*	*ethnicity*	*nonresident_alien*	Description
Cluster 0	40.7%	White	0%	White, very high percent without internet
Cluster 1	8.7%	White	0%	White, very low percent without internet
Cluster 2	19.8%	Black/African American	1%	Black students, average percent without internet
Cluster 3	16.7%	White	0%	White, average percent without internet
Cluster 4	26.9%	White	0%	White, high percent without internet

6 Cluster Creation and Performance – C165 Integrated Physical Science

6.1 Clustering Method

Just as we had done for the biochemistry course, each of the physical science datasets was uploaded to WEKA Explorer and clustered using KMeans method into five different clusters.

We again recorded the following data points (number of clusters, minimum and maximum percent of total student population, spread in percent of total population, minimum and maximum course completion rate, and spread in course completion rate), and used these to evaluate the relative value of the different datasets for cluster creation.

Table 11. Category Performance – C165

Category	Min % of Population	Max % of Population	Spread in Population	Min % Completion	Max % Completion	Spread in Completion
Activity - Week 1	13.3%	27.6%	14.4%	70.8%	88.6%	17.8%
Activity- Week 2	16.3%	23.1%	6.7%	58.3%	90.7%	32.5%
Activity - Week 3	11.2%	25.6%	14.4%	57.4%	93.7%	36.2%
Demographic	7.1%	32.2%	25.1%	67.0%	81.5%	14.5%

What we see across the activity categories is a sharp increase in goodness of fit from week 1 to week 2 (with the outcome differentiation growing from 17.8% to 32.5%) with a much smaller increase between weeks 2 and 3 (from 32.5% to 36.2%).

The demographic category is slightly behind activity measured at week 1 (14.5% and 17.8% respectively).

As early as seven days into a six-month term – less than 5% of the time of the term – activity data shows more predictive power than demographic data. This is even after using attribute selection to identify the most valuable of the demographic attributes.

6.2 Differences by Student College in Outcome Variable – C165

Because the Integrated Physical Science (C165) course fills a general education requirement, it is taken by students from three of the four colleges (I.T., Business, and Teachers). Health Professions students are extremely rare in this specific course, having their science requirement met by several of the courses in their program. We started by comparing the outcome differentiation (% difference in course completion) across the three clusters.

It is important to note that students need not be in the same number cluster across the three weeks. The clusters are based on student activity through the end of the respective week, and students frequently had different activity patterns week over week throughout the early part of the term. This would be captured in a change for the student cluster assignment.

Table 12. Cluster Outcome Differentiation by College and Activity Cluster

Category	Week 1 Avg Completion %	Week 2 Avg. Completion %	Week 3 Avg. Completion %
Business –cluster 0	65.2%	91.4%	70.9%
Business – cluster 1	72.1%	80.2%	93.2%
Business – cluster 2	88.9%	63.9%	87.8%
Business – cluster 3	82.5%	77.2%	80.0%
Business – cluster 4	78.3%	68.9%	62.7%
IT – cluster 0	68.4%	93.9%	79.1%
IT – cluster 1	80.5%	88.9%	95.8%
IT – cluster 2	92.5%	67.7%	91.0%
IT – cluster 3	89.2%	84.1%	88.0%
IT – cluster 4	86.8%	78.4%	67.7%
Teachers – cluster 0	58.8%	89.6%	63.9%
Teachers – cluster 1	68.2%	77.9%	92.5%
Teachers – cluster 2	87.4%	57.4%	84.4%
Teachers – cluster 3	79.8%	70.4%	76.4%
Teachers – cluster 4	74.5%	62.0%	56.6%

We see here that the specific cluster-week combination varies for all colleges between the three weeks of interest. As mentioned above, students did not remain in the same number cluster across the three weeks, which accounts for the differences we say in the outcome variable (completing the course) as we moved from week to week in each college-cluster combination.

6.3 Differences by Student College in Population Discrimination – C165

Because the Integrated Physical Science (C165) course fills a general education requirement, it is taken by students from three of the four colleges (I.T., Business, and Teachers). Health Professions students are extremely rare in this specific course, having their science requirement met by several of the courses in their program. After comparing the outcome differentiation across the clusters/colleges, we move on to compare the cluster differentiation (% of total population).

Table 13. Cluster Population Differentiation by College and Activity Cluster

Category	Week 1 Avg Population %	Week 2 Avg. Population %	Week 3 Avg. Population %	Total Avg. Population %
Business –cluster 0	22.0%	18.0%	25.0%	21.7%
Business – cluster 1	28.1%	21.0%	9.8%	19.6%
Business – cluster 2	18.5%	22.7%	15.6%	19.0%
Business – cluster 3	12.0%	20.9%	25.9%	19.6%
Business – cluster 4	19.4%	17.4%	23.6%	20.1%
IT – cluster 0	16.1%	23.0%	22.5%	20.6%
IT – cluster 1	26.8%	24.1%	12.4%	21.1%
IT – cluster 2	23.4%	18.0%	19.9%	20.4%
IT – cluster 3	14.1%	19.8%	26.0%	20.0%
IT – cluster 4	19.6%	15.1%	19.1%	17.9%
Teachers – cluster 0	17.8%	22.9%	23.2%	21.3%
Teachers – cluster 1	27.4%	24.8%	12.4%	21.5%
Teachers – cluster 2	22.5%	17.3%	20.5%	20.1%
Teachers – cluster 3	14.4%	18.7%	26.5%	19.9%
Teachers – cluster 4	18.0%	16.2%	17.5%	17.2%

Overall, we see fairly consistent population differentiation. At any given week, some clusters will have more or less of the overall college population, but when we average for that cluster-college across all three weeks, our values are between 17.2% and 21.7% of the total college population. Our smallest individual population proportion is for the Business College students – cluster 1 at week 3 with 9.8% of the total population for that week.

7 Discussion of Findings – Recommendations for Adaptive Systems

7.1 What attributes matter to student success?

We are aware that many student attributes – and even the phrase "students like you" – have a heavy connotation, and, in many instances, legal protection. As we started this study, we hoped to find a combination of student attributes with enough differentiating power to outweigh the negative connotations. We intentionally chose difficult courses – so there would be enough variation in student outcomes to make the analysis more meaningful. We quantitatively measured the incremental value of the different categories of attributes to create a framework of selected attributes with data-driven support. We intended to use these to recommend enhancement and meaningful learning guidance for students – to show that it would be worthwhile to include these attributes in adaptive systems because the benefit of the attributes would outweigh the connotation and risk of including them.

We were surprised to find that the attributes with the most value were also those with the least cultural sensitivity. We found behavior to greatly exceed the value of both the readiness assessment and demographic attribute categories (although both readiness and demographic attributes did add to the value of behavior when the categories were combined). Especially when we take into account the spread in population from the individual categories to the combined category, student activity attributes as a standalone category have the greatest value. An increase in discrimination from 35.8% (activity alone) to 37.8% (activity & readiness) or 41.8% (activity and demographics) reduces the population discrimination from 29.9% to 7.9% and 11.2% respectively. The relative cost of adding one of the other categories to activity is to greatly reduce the size of the clusters – to the point that smaller datasets (C785 datasets included over 45,000 students) might not get enough meaningful clusters to proceed.

7.2 What should designers of adaptive learning resources do with this?

There are several reasons why these results – though not what we expected – are extremely encouraging to designers, educators, and students alike. We find that the most differentiating characteristics – with the biggest difference in student success – are those with the most potential for change. Student behavior persisted as the best way to identify ultimately successful students – with large gaps in outcomes showing for students alike in all but their behavior.

For designers and developers of learning resources, the value of activity is promising because these are the attributes which are the easiest to gather from learning resource behavior – without relying on systems sharing protected data about students. Organizations that host resources for students can see the activity data without requiring any additional (and protected) data about the students. Adaptive systems can be designed and built based on activity data alone, which is the easiest and safest to collect for these systems.

7.3 What should instructors and students do with this?

For instructors and students this is encouraging because – while changing behavior is no small feat, it is possible, and both resources and instructional plans can be built accordingly. Based on our findings in this study and additional previous research on the Doer Effect [1, 2, 3, 4], we recommend designing courses with frequent opportunities for active practice. Additionally, researchers in earlier studies [5] have found that student participation in these practice opportunities is helped by awarding some points for their completion. This gives both students and instructors reasons and ways to motivate changes in activity.

8 Next Steps and Research Questions

8.1 Exploratory Work

The work we discuss in this paper is very early and exploratory. We did not find what we expected but we are very encouraged by what we did find. As we have discussed activity has the most power in differentiating students and is also the easiest thing to guide and impact.

Possible directions for future analysis include evaluating student outcome not as a binary P/F variable, but instead using the continuous variable of percentage scores – perhaps restricting to individual competencies or learning objectives.

Additionally, we are already planning to include additional variables as part of the day 0 data set – including engineered features from their enrollment timeline, diagnostic data specific to the course content, and others.

We have also discussed evaluation of the clustering attributes as a single dataset used for attribute selection, rather than selecting/clustering attributes for each category individually.

Finally, we find ourselves asking at what point the growth of activity data in value taper off? We have seen a very significant growth between one and two weeks, with continued – though less significant – growth between two and three weeks. At some point, any incremental gain in goodness of fit is likely to be outweighed by the relative cost of waiting that much longer to adapt, something which requires additional analysis or understanding of the value of adaptation at different points in the term. In order to support this discussion with data, an objective for future stages of this research includes understanding where that point is found.

References

1. Koedinger, K., Kim, J., Jia, J., McLaughlin, E., & Bier, N. (2015). Learning is not a spectator sport: doing is better than watching for learning from a MOOC. In: Learning at Scale, pp.111-120. Vancouver, Canada. http://dx.doi.org/10.1145/2724660.2724681
2. Koedinger, K., Kim, J., Jia, J., McLaughlin, E., & Bier, N. (2015). Is the doer effect a causal relationship? How can we tell and why it's important. In: Learning Analytics and Knowledge. Edinburgh, United Kingdom. http://dx.doi.org/10.1145/2883851.2883957

3. Koedinger, K. R., Scheines, R., & Schaldenbrand, P. (2018). Is the doer effect robust across multiple data sets? Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, 369–375.
4. Olsen, J., Johnson, B. (2019). Deeper collaborations: a finding that may have gone unnoticed. In: IMS Global Learning Impact Leadership Institute. San Diego, CA.
5. Edgcomb, A., Vahid, F. (2015). How many points should be awarded for interactive textbook reading assignments? In: IEEE Frontiers in Education Conference (FIE). El Paso, TX. http://dx.doi.org/10.1109/FIE.2015.7344350.