Office of the Provost
INSTITUTIONAL EFFECTIVENESS AND PLANNING
George Mason University

Analysis of the Knack Tutoring Platform
Fall 2025

Please note: the following discussion focuses on descriptive statistics and associations related to course outcomes and Knack tutoring use. There is no assertion of causality.

SUMMARY

Since Fall 2024, when George Mason began partnering with a free peer-to-peer tutoring platform called Knack, the question of whether the platform is helping students succeed has become of interest to those on campus. An initial review of the data indicated that courses with Knack had higher average grades in Fall 2024, when Knack was available, compared to prior years. However, it was unclear if that improvement can be attributed to Knack use.

To assess the effectiveness of the program beyond course-level averages, three matching schemes were used to compare the outcomes of Knack users to similar students who did not use Knack. A Difference-in-Differences model was also employed to assess whether there was a relationship between improved course-level outcomes and the introduction of the platform. Across the various designs, there was a consistent lack of evidence that Knack was associated with improved grades and course outcomes.

More detail is provided below.

INTRODUCTION

In Fall 2024, George Mason began partnering with Knack, a free peer-to-peer tutoring platform, in which students can access tutoring for select courses each semester. The tutors are current George Mason students who succeeded in the same course. To assess the effectiveness of the program, OIEP was asked to analyze the success of Knack users in the courses for which they received tutoring.

MATCHED ANALYSIS

Methods

Not all students have the same likelihood of needing the type of assistance Knack provides. To account for this and only compare Knack users to those similarly likely to need tutoring, a nearest neighbor propensity score matching was employed and was based on a range of demographic and academic variables: prior GPA, race, Pell eligibility, student level (e.g., freshman), age, first-generation

status, gender, class load (e.g., part time), and in-state status. For new students, their high school (HS) GPA was used to understand their previous achievement instead of their college-level cumulative GPA. To further ensure the validity of matches, students had to be in the same course as their match and their GPA had to be from the same source (whether cumulative college GPA or high school GPA).

Table 1. Means of Original and Matched Data Set, by Knack Tutoring Use

	Unmatched			Matched		
	Comparison	Knack	p-Value	Comparison	Knack	p-Value
Average GPA	2.98	2.77	0.0029	2.75	2.77	0.8125
Percent with Only a HS GPA	0.2747	0.2833	0.722	0.2833	0.2833	1
Average Age	20.3153	21.0556	0.0028	20.5694	21.0556	0.1446
Percent Nonresident Alien (NRA)	0.0929	0.0528	0.0011	0.0583	0.0528	0.7453
Percent Part Time	0.0871	0.1222	0.046	0.1194	0.1222	0.9091
Percent Freshmen	0.2762	0.2889	0.6029	0.2861	0.2889	0.9345
Percent In State	0.8271	0.8056	0.3126	0.8056	0.8056	1
Percent Female	0.4806	0.6806	0	0.7083	0.6806	0.4192
Percent Pell-eligible	0.3326	0.3861	0.0416	0.375	0.3861	0.7592
Percent First Generation	0.25	0.26	0.5854	0.25	0.26	0.732
Total Observations	8726	360		360	360	

Table 1 compares the student characteristics of those who used Knack tutoring to those in the same semester who did not use Knack, before and after matching. The unmatched dataset consists of all students who were in courses where Knack was available. The matched dataset is a subset of those students, keeping one classmate per Knack user who is closest in relevant demographics. Independent samples t-tests were run to compare the means. p-Values for these tests are shown in the table. In the original, unmatched set, several significant differences were observed: Knack users had lower previous GPAs (2.77 vs. 2.98), were older (21.06 vs. 20.32 years), and were less likely to be international students (5% vs. 9%). After matching, these differences were reduced to statistically insignificant levels. The higher p-values for the matched dataset are partially due to the lower sample size. However, the matched group shows greater balance and noticeable convergence between the comparison and Knack groups. This is especially evident in the percent of students who are female, where the difference between groups before matching was roughly 20 percentage points. After matching, this difference decreased to 2.8 percentage points. This suggests that the matching did enhance the comparability of the Knack and control groups, and the higher p-values in the matched dataset cannot be solely attributed to the lower number of observations.

With increased balance between the group who used Knack tutoring and the group that did not, the matched dataset was then used to test whether using Knack was associated with higher grades. A regression of course grades, which range from zero to four with four being the best grade, was conducted. Fixed effects were included for each pair. The regression model is shown in Formula 1:

$$(1)\quad \text{grdpts}_{ij} = \beta_0 + \beta_1 \cdot \text{knack\_indicator}_{\,i} + \beta_2 \cdot \text{GPA}_i + \beta_3 \cdot \text{Race}_i + \beta_4 \cdot \text{Non\_Trad\_Age}_{\,i} + \beta_5 \cdot \text{Student\_Type}_i + \beta_6 \cdot \text{Pell}_i + \beta_7 \cdot \text{Part\_Time\_Load}_{\,i} + \beta_8 \cdot \text{First\_Gen}_i + \beta_9 \cdot \text{Sex}_i + \beta_{10} \cdot \text{Domicile}_i + \gamma_j \cdot \text{Pair\_ID}_{\,j} + \varepsilon_i$$

where: grdpts_{ij} is the course grade points for student i in pairing j. β_1 is the coefficient for $\text{knack\_indicator}_{\,i}$, which is the predictor variable of interest. $\text{knack\_indicator}_{\,i}$ is specified in two

ways to fully understand the effects of Knack tutoring. A binary variable indicating which students used Knack is analyzed, then the number of Knack sessions attended by a student is substituted to further understand if more Knack sessions were associated with better outcomes. β_2 to β_{10} are coefficients representing the effects of various student-level control variables. γ_j are the coefficients for matched pair fixed effects, indexed by j.

Regression Results

The matched regression shows a significant, negative association between Knack tutoring and grades, with a coefficient of -0.23, which is nearly the distance between a B- and a C+, for example. This does not mean that Knack caused lower grades; it is more likely that Knack students were already struggling in the class and Knack was not able to fully mitigate those challenges. Not all students who used Knack had the same tutoring intensity, with Knack students attending anywhere from 1 to 28 sessions. To account for this, the number of sessions, rather than a dummy variable for Knack usage, was substituted as the variable of interest. The matched regression showed no significant association between the number of Knack sessions and grades. Together, it seems that Knack usage may be associated with lower grades, but since the comparison group also had access to Knack, this design cannot account for any unmeasured factors, such as difficulty with specific course materials, that caused some students to choose to use Knack while others did not.

Table 2. Matched Regression Analyses of Course Grades for Students in Fall 2024

Predictor	Coefficient (SE) for Knack Use Model	Coefficient (SE) for Knack Sessions Model
Knack Use	-0.228* (0.088)	
Knack Sessions		-0.009 (0.015)
GPA	1.34*** (0.331)	1.267*** (0.333)
NRA	0.474 (0.428)	0.493 (0.433)
Freshmen	0.149 (1.013)	0.107 (1.023)
Out of State	-0.407 (0.255)	-0.415 (0.258)
Fixed Effect, Pair 5	-1.996 (1.828)	-2.378 (1.839)
Fixed Effect, Pair 4	-3.213 · (1.862)	-3.124 · (1.893)
Fixed Effect, Pair 3	-4.765* (1.88)	-4.811* (1.898)

Observations: 708,
R-Squared: 0.64, 0.633
Adj R-Squared: 0.214, 0.198
Significance codes: '***' Pr(>|t|) < 0.001; '**' Pr(>|t|) < 0.01; '*' Pr(>|t|) < 0.05; '.' Pr(>|t|) < 0.1

MATCHED ANALYSIS, ALTERNATE SPECIFICATIONS

Methods

While matching can help ensure that students in a control and treatment group are balanced on observable characteristics, there are unobserved characteristics that make some students more likely to seek Knack tutoring than their peers who may not have sought tutoring at all. To account for this, the sample was limited only to students who accessed a University Life (UL) program that included academic supports. They are: CSS Peer mentoring, FGC Student Appointments, GSL Academic and Professional Success, HRL Learning Communities, OIPS Co-Curricular Learning, LS Learning Support Services, SSC Patriot Experience, SSC Student Success Coaching, UCS Career Development. Although the ways in which students interacted with each program is unknown, these students sought support,

as Knack students did, suggesting that both have some unmeasured tenacity or need for assistance. Using the same process as was used for the full sample, Knack students were matched to other students who sought UL support.

Table 3 compares the student characteristics of those who used Knack tutoring to those who used other UL supports, before and after matching. In the unmatched set, several significant differences were observed, notably that Knack users were older, less likely to be freshmen, and more likely to be part-time. The differences were more significant in this limited sample than in the full sample, which is shown in Table 1. This suggests that Knack students, on the whole, are different types of students than their counterparts who use UL supports. After matching, though, these differences were generally reduced to statistically insignificant levels, except for age. The gap between the group averages dropped from 1.6 years to 0.9 years after matching, but this remained a statistically significant difference. This raises some questions about the comparability of the two groups, though this will be partially accounted for by the inclusion of age, and other student-level factors, as controls in the regression. Nevertheless, regression results should be interpreted with caution.

Table 3. Means of Original and Matched Data Set, Limited to Those who Used Knack or UL services, by Knack Tutoring Use

	Unmatched			Matched		
	Comparison	Knack	p-Value	Comparison	Knack	p-Value
Average GPA	3.15	2.77	0	2.76	2.77	0.935
Percent with Only a HS GPA	0.4096	0.2833	0	0.2833	0.2833	1
Average Age	19.5009	21.0556	0	20.1528	21.0556	0.0016
Percent Nonresident Alien (NRA)	0.0866	0.0528	0.0072	0.0722	0.0528	0.2818
Percent Part Time	0.0404	0.1222	0	0.1167	0.1222	0.8185
Percent Freshmen	0.4132	0.2889	0	0.2861	0.2889	0.9345
Percent In State	0.7982	0.8056	0.7369	0.7861	0.8056	0.5182
Percent Female	0.5263	0.6806	0	0.6972	0.6806	0.6297
Percent Pell-eligible	0.3355	0.3861	0.0583	0.3889	0.3861	0.9391
Percent First Generation	0.22	0.26	0.1199	0.26	0.26	0.9321
Total Observations	4456	360		360	360	

Given the differences between the group who used UL services and Knack students, there remained questions about whether contemporaneous students provide the best counterfactual for Knack users. Comparing Knack students to those who used UL services can account for unmeasured willingness to seek help. However, students most similar to Knack users also had the opportunity to use Knack but for an unknown reason chose not to use it. On the other hand, students enrolled in Fall 2023 did not have access to Knack but some would have been likely to use the platform had it been available, making them a useful comparison group. With this distinction in mind, Knack users in Fall 2024 were matched to students in Fall 2023 who were enrolled in the same class. Looking across years introduces new potential confounders but provides a complement to the within-semester analyses.

Table 4. Means of Original and Matched Data Sets, comparing Knack Users to Students in Fall 2023 taking the Same Courses, by Knack Tutoring Use

	Unmatched			Matched		
	Comparison	Knack	p-Value	Comparison	Knack	p-Value
Average GPA	2.97	2.77	0.0041	2.77	2.77	0.9945
Percent with Only a HS GPA	0.293	0.2833	0.6898	0.2833	0.2833	1
Average Age	20.3692	21.0556	0.0056	20.9972	21.0556	0.8658
Percent Nonresident Alien (NRA)	0.0838	0.0528	0.0112	0.0639	0.0528	0.5254
Percent Part Time	0.0914	0.1222	0.0799	0.0972	0.1222	0.2838
Percent Freshmen	0.2933	0.2889	0.8578	0.2889	0.2889	1
Percent In State	0.8259	0.8056	0.3402	0.7944	0.8056	0.7099
Percent Female	0.5078	0.6806	0	0.6722	0.6806	0.8114
Percent Pell-eligible	0.3046	0.3861	0.002	0.4278	0.3861	0.2558
Percent First Generation	0.23	0.26	0.2972	0.28	0.26	0.5565
Total Observations	8569	360		360	360	

Table 4 compares the student characteristics of those who used Knack tutoring to those in the same courses in Fall 2023, before and after matching. In the unmatched set, as with the within-semester comparison, several significant differences were observed, although these differences were less stark than in the unmatched set comparing Knack students to those who accessed UL supports. After matching, these differences were reduced to statistically insignificant levels. The impact of Knack in both matched datasets was analyzed using Formula 1.

Regression Results

Using the matched data comparing Knack students to their peers who used UL services in Fall 2024, the regression shows that Knack students had significantly lower grades than their counterparts, with a coefficient of -0.21. This is almost the difference between a B- and a C+, for example. Again, this is not to say that Knack usage caused lower grades; it is more likely that Knack students were struggling in the class more than their peers and Knack was not able to fully mitigate those challenges. Just over 68% of Knack students also sought other UL supports. For example, 12% were in a learning community and 33% sought a success coach appointment. Together, this paints a picture of students who needed more assistance than their peers, above and beyond what could be accounted for using the control variables.

On the other hand, when comparing Knack students to similar students in Fall 2023, the impact of Knack on grades was not statistically significant. If there was something fundamentally different about Knack students and their peers in the same semester, then the cross-year regression may provide a better comparison. If so, the insignificant result suggests that Knack was not associated with a change in grades for users. Either way, none of the matched regressions showed a relationship between improved grades and Knack use.

Table 5. Matched Regression Analyses of Course Grades, Using Alternate Comparison Populations

Predictor	Coefficient (SE) for UL Comparison Model	Coefficient (SE) for Fall 2023 Comparison Model
Knack Use	-0.213* (0.087)	0.074 (0.087)
GPA	1.225*** (0.355)	1.108** (0.337)
NRA	0.946* (0.366)	0.633 (0.502)
Freshmen	-1.936 (1.291)	1.453 (1.955)
Out of State	-0.316 (0.214)	-0.647 · (0.363)
Fixed Effect, Pair 5	0.843 (1.183)	2.714 (1.86)
Fixed Effect, Pair 4	-0.16 (1.242)	1.124 (1.448)
Fixed Effect, Pair 3	-1.32 (1.168)	-3.086 (2.232)

Observations: 708,
R-Squared: 0.67, 0.654
Adj R-Squared: 0.281, 0.247
Significance codes: '***' Pr(>|t|) < 0.001; '**' Pr(>|t|) < 0.01; '*' Pr(>|t|) < 0.05; '.' Pr(>|t|) < 0.1

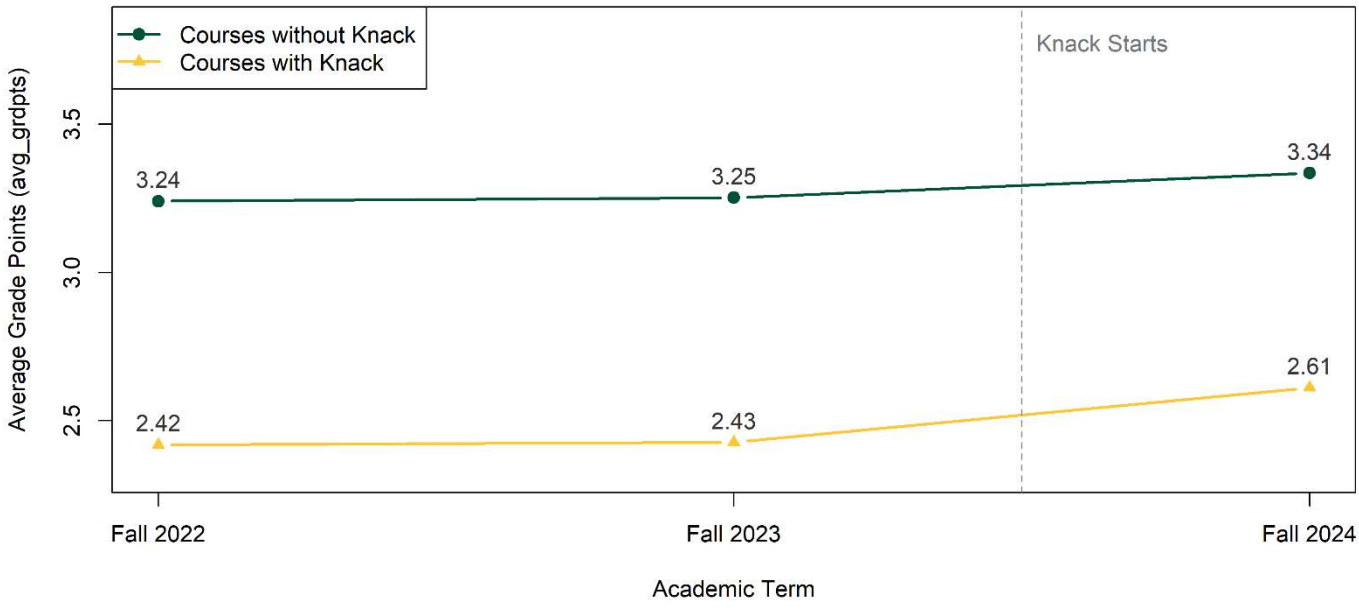
Note: When the number of Knack sessions was substituted as the variable of interest, the coefficient for Knack sessions was statistically insignificant for both matching schemes.

DIFFERENCE-IN-DIFFERENCES MODEL

Methods

The unobserved characteristics of Knack users, compared to their counterparts, remain a concern, especially if those characteristics are associated with lower grades and are biasing the effect of Knack downward. To address this, a difference-in-differences (DiD) analysis was completed. This is a helpful complement to the matched analysis, if the assumption holds that there were students in Fall 2022 and Fall 2023 with similar unobserved traits as those in Fall 2024, but who did not have access to Knack services. This is in addition to the underlying assumption of DiD analyses, that there are parallel trends between the treatment and control groups in the pre-period. As shown in Figure 1, this seems to be the case.

Figure 1. Average Grade Points Over Time, by Whether Knack was Offered in that Course in Fall 2024.



To prepare the dataset for the DiD analysis, student-level records were compiled across three terms: Fall 2022, Fall 2023, and Fall 2024. The records were restricted to relevant undergraduate courses with nonmissing course grades. Averages were then computed for each course and term to aggregate student-level measures to the term-by-course level. These included the outcome variables (i.e., grade points and DFW rates) and a range of control variables (e.g., cumulative GPA, percent who are freshmen, percent who are NRA students, the number of students taking the course that semester). Then, the treatment period indicator was created to distinguish the pre-Knack terms, Fall 2022 and 2023, and the post-Knack term, Fall 2024. Similarly, courses in which students used Knack tutoring were flagged as treated courses. Finally, an interaction term between the treatment flag and the treatment period indicator was generated to capture the DiD effect.

A preliminary review of the data, as shown in Figure 1, indicates that students in courses with Knack did have higher average grades in Fall 2024, when Knack was available, compared to prior years. Students in courses where Knack was available had a 2.43 average in Fall 2023 and a 2.61 average in Fall 2024, representing a 0.184 increase. Other courses also saw an increase in grades from Fall 2023 to Fall 2024, but it was smaller at 0.083.

To check if the increase in grades can be attributed to the availability of Knack tutoring, a course-level DiD model was completed. It is specified in Formula 2:

$$\begin{aligned}\mathrm{Outcome}_{tj} = {} & \beta_0 + \beta_1 \cdot \mathrm{Term}_t + \beta_2 \cdot \mathrm{Knack\_Course}_j + \beta_3 \cdot \mathrm{DiD}_{tj} + \beta_4 \cdot \mathrm{avg\_gpa}_{tj} + \beta_5 \cdot \mathrm{pct\_URM}_{tj} \\ & + \beta_6 \cdot \mathrm{pct\_NRA}_{tj} + \beta_7 \cdot \mathrm{pct\_Pell\_eligible}_{tj} + \beta_8 \cdot \mathrm{avg\_age}_{tj} + \beta_9 \cdot \mathrm{pct\_first\_gen}_{tj} + \beta_{10} \\ & \cdot \mathrm{pct\_female}_{tj} + \beta_{11} \cdot \mathrm{pct\_in\_state}_{tj} + \beta_{12} \cdot \mathrm{student\_count}_{tj} + \beta_{13} \cdot \mathrm{pct\_part\_time}_{tj} + \varepsilon_{tj}\end{aligned}$$

where $\mathrm{Outcome}_{tj}$ is specified in two ways: first as the mean course grade for course j in term t and second as the DFW rate for course j in term t. DFW rates are added as a dependent variable for the DiD analysis only because it is a course-level metric and provides additional information about Knack's impacts on overall course outcomes. The coefficient on the interaction term, DiD_{tj}, represents the difference-in-differences effect estimating the impact of Knack tutoring on course outcomes. The model also includes a term fixed effect (Term_t). The remaining variables are course-level controls, including mean cumulative GPA, percent of students in the course who are freshmen, percent who are NRA students, percent who are underrepresented minority (URM) students, percent who are female students, average student age, percent who are part time, and the number of students taking the course.

Regression Results

For both average course grades and DFW rates, the coefficient on the DiD indicator is statistically insignificant. This suggests that Knack tutoring was not associated with improved course-level outcomes. Given that no course had more than 8% of the students using Knack in Fall 2024, it is maybe unsurprising that it did not have significant impacts on overall outcomes. Nevertheless, average grades in Knack courses did increase in Fall 2024. As Knack continues to be available, additional treatment years will help determine whether that trend continues or is merely coincidental.

Table 6. DiD Analysis of Course Grades and DFW Rates, Selected Regression Coefficients

Predictor	Coefficient (SE) for the Average Grades Model	Coefficient (SE) for the DFW Model
Post-Knack Period (Fall 2024)	0.105*** (0.017)	-0.016*** (0.003)
Knack Course	-0.613*** (0.078)	0.113*** (0.015)
DiD, Fall 2024 Knack Courses	0.117 (0.131)	-0.033 (0.026)
Average GPA	0.289*** (0.016)	-0.049*** (0.003)
Percent NRA	0.18* (0.075)	-0.041** (0.015)
Percent Freshmen	0.629*** (0.068)	-0.071*** (0.013)
Percent In-State	-0.016 (0.047)	0.005 (0.009)

Observations: 4685,
R-Squared: 0.166, 0.131,
Adj R-Squared: 0.164, 0.128
Significance codes: '***' Pr(>|t|) < 0.001; '**' Pr(>|t|) < 0.01; '*' Pr(>|t|) < 0.05; '.' Pr(>|t|) < 0.1

CONCLUSION

None of the matched regressions or DiD analyses showed a significant relationship between improved academic outcomes and Knack use. Of course, there are limitations to each design that warrant consideration. Overall, students and courses were not randomly selected into the Knack program, so there will always be identification challenges when assessing its impact. Across the various designs, though, there is a consistent lack of evidence that Knack is associated with improved grades and course outcomes.

Going forward, data from additional semesters can be added to further assess Knack's impacts. This is especially true for the DiD analysis where additional years may reveal Knack-related trends that are currently hard to discern. In addition, further work can be done to identify student characteristics that help account for currently unmeasured differences between Knack users and their peers. For example, future analyses could compare Knack students to peers who sought other academic supports for the same class, rather than to any student who accessed UL services with an academic component. That may require additional data collection over what is currently available to OIEP. These kinds of additional analyses can provide insights into the value of Knack on campus.