Kurdish Studies
Feb, 2021
Volume: 09, No: 01, pp 206-216
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

**DOI:** 10.67145/ks.v9i1.4131

# The Relationship Between Parental Involvement and Academic Performance in Basic and Secondary Education Students

**Luis Eduardo Muñoz Guerrero**[1]
[1]Universidad Tecnológica de Pereira. Email: lemunozg@utp.edu.co. ORCID ID: https://orcid.org/0000-0002-9414-6187

**Abstract**
**Parental involvement is widely regarded as one of the most consistent correlates of student academic success, yet many of the datasets used to study this relationship are either proprietary, longitudinal, or too large to be feasible for a single, unfunded researcher to analyze. This study examines the association between two observable proxies of parental involvement — whether a parent answered the school's electronic survey and the parent's reported satisfaction with the school — and the academic performance level of 480 students, using the publicly available xAPI-Edu-Data dataset (Amrieh, Hamtini, & Aljarah, 2016). A quantitative, cross-sectional, descriptive-correlational design was used, combining descriptive statistics, cross-tabulation, and chi-square tests of independence. Results confirmed this expectation: students whose parents answered the school survey or reported satisfaction with the school were significantly more likely to fall into the High performance category, $\chi^2_{2,N=480}=95.36, p<.001$, Cramér's V=.45 for survey response, and $\chi^2_{2,N=480}=68.47, p<.001$, Cramér's V=.38 for reported satisfaction. These findings provide low-cost, replicable evidence on the role of family engagement in school performance, of practical relevance to teachers, school counselors, and educational administrators interested in low-resource, evidence-based family-involvement strategies.**

***Keywords:*** parental involvement; academic performance; educational data mining; xAPI-Edu-Data; family engagement; secondary education

## 1. Introduction

The idea that parents shape how well their children do in school is intuitively appealing and has motivated decades of educational research, yet the empirical evidence behind it is more nuanced than common sense suggests. Large-scale syntheses of the literature confirm that parental involvement is associated with better academic outcomes, but the size of that association varies considerably depending on how involvement is defined, at what educational level it is measured, and which specific behaviors are counted as "involvement" (Fan & Chen, 2001; Jeynes, 2007). This variability suggests that the relationship, while real, is not uniform, and that it deserves continued, context-specific investigation rather than being treated as a settled fact.

Historically, research on parental involvement required researchers to design and administer their own surveys, secure institutional access to schools, and manage the ethical and logistical demands of working directly with minors and their families — a substantial barrier for an independent researcher without external funding or institutional backing. The growth of educational data mining and the increasing availability of anonymized, open-access datasets derived from learning management systems has begun to lower that barrier. Learning management systems routinely capture indicators that were previously invisible to researchers, such as whether a parent responded to a school survey or how satisfied a parent reports being with the institution, alongside detailed behavioral and academic records for each student. These secondary, already-collected data sources make it possible to study family-engagement questions on a small scale, at no financial cost, and without direct contact with human subjects beyond the original, already-approved data collection.

This study takes advantage of one such dataset — the xAPI-Edu-Data set, originally compiled by Amrieh, Hamtini, and Aljarah (2016) from a Jordanian e-learning platform — to examine, on a small and manageable scale, whether two simple, observable proxies of parental involvement (survey response and reported school satisfaction) are associated with students' final academic performance level. Unlike the original study, which focused on predictive modeling of performance using ensemble machine-learning methods, the present study adopts a simpler, explanatory, descriptive-correlational approach that a single researcher can complete without specialized computational resources.

The specific research question guiding this study is: Is there a statistically significant association between parental involvement, as proxied by (a) whether a parent answered the school survey and (b) the parent's reported satisfaction with the school, and students' academic performance level (Low, Medium, or High)? Two corresponding hypotheses are proposed. The first (H1a) states that students whose parents answered the school survey will show a significantly different distribution of performance levels compared to students whose parents did not. The second (H1b) states that students whose parents report being satisfied with the school will show a significantly different distribution of performance levels compared to students whose parents report dissatisfaction. The corresponding null hypotheses (H0a and H0b) state that no such association exists in either case.

This study contributes to the literature in three modest but concrete ways. First, it offers a low-cost, fully replicable analysis pipeline — built entirely on free, open-access data and open-source statistical tools — that other single-researcher or student projects can adapt to similar questions. Second, it extends the use of the xAPI-Edu-Data set, which has been used

primarily for predictive modeling, toward a more traditional explanatory framework centered on family engagement rather than algorithmic prediction. Third, by keeping the design intentionally simple — descriptive statistics, cross-tabulation, and chi-square testing — the study aims to produce results that are transparent and interpretable to non-specialist educational stakeholders, such as classroom teachers and school administrators, who are often the intended audience for applied educational research of this kind.

**2. Literature Review**

Parental involvement has been studied under many different operational definitions, ranging from direct homework assistance to broader forms of engagement such as attending school events, communicating with teachers, or simply holding high expectations for a child's academic future. Epstein's (2001) influential framework organizes these behaviors into six overlapping types: parenting, communicating, volunteering, learning at home, decision-making, and collaborating with the community. Of these, communicating — the exchange of information between families and schools, including responses to school surveys and expressed satisfaction with the institution — and the broader construct of family-school partnership are the most directly relevant to the proxies available in the dataset used in the present study.

The most influential quantitative synthesis of this literature remains Fan and Chen's (2001) meta-analysis, which pooled twenty-five empirical studies published between 1982 and 1997. The authors reported a small-to-moderate but practically meaningful overall association between parental involvement and academic achievement, and noted that the strength of this association depended heavily on the specific type of involvement measured: parental expectations and aspirations for their children's education showed the strongest relationship with achievement, while more passive forms of involvement, such as general home supervision, showed the weakest. This distinction is important for the present study, since "parent answered the school survey" and "reported satisfaction with the school" are best understood as indicators of engaged, communicative involvement rather than direct academic supervision, and might therefore be expected to show a moderate, rather than a strong, association with performance.

This pattern is not unique to Fan and Chen's (2001) synthesis. In a later meta-analysis of 37 studies conducted between 2000 and 2013, Castro et al. (2015) similarly found that the strongest associations with achievement emerged when families held high academic expectations and maintained ongoing communication about school activities, rather than when involvement was limited to general supervision.

Wilder (2014), synthesizing nine prior meta-analyses into a single meta-synthesis, reached a similar conclusion: parental involvement operationalized as expectations and aspirations tends to show a stronger relationship with achievement than involvement operationalized as more passive or narrowly communicative behaviors, a pattern reported as consistent across the different grade levels and ethnicities represented in the underlying studies. Because ParentAnsweringSurvey and ParentschoolSatisfaction fall closer to the passive, communicative end of this spectrum than to the expectation end, this converging evidence further supports the expectation that any association observed in the present study is likely to be modest rather than large.

Subsequent meta-analyses extended these findings to different educational levels and contexts. Jeynes (2007) synthesized fifty-two studies conducted between 1972 and 2002, specifically examining urban secondary school students, and found a meaningful, statistically significant overall effect of parental involvement on academic achievement at the secondary level — a finding of particular relevance here, since the xAPI-Edu-Data dataset includes students from lower and middle school grade levels comparable to secondary education in many international contexts. Jeynes further distinguished between subtle, emotionally supportive forms of parental involvement (such as general encouragement and communication) and more instrumental forms (such as strict monitoring or attendance at school functions), reporting that the subtler forms were often at least as strongly associated with achievement as the more visible, instrumental ones. This distinction reinforces the plausibility of studying communicative proxies, such as survey response and reported satisfaction, as legitimate indicators of family engagement.

The developmental stage of the student sample is also relevant to how these proxies should be interpreted. Hill and Tyson's (2009) meta-analysis of 50 studies conducted specifically at the middle-school level found that most forms of parental involvement were positively associated with achievement, with the notable exception of direct help with homework, and that involvement reflecting academic socialization — communicating expectations and the importance of school — had the strongest positive association of all the types examined.

A companion meta-analysis by Jeynes (2005), covering elementary-level students, extended this same line of synthesis to a younger age group. Kim and Hill (2015) further showed, across kindergarten-to-12th-grade studies, that the relationship between involvement and achievement was similarly strong regardless of whether the involved parent was the father or the mother, which is relevant here because the xAPI-Edu-Data set records which parent (Relation) answered the survey without this distinction being examined further in the present design. Finally, Pomerantz, Moorman, and Litwack (2007) argued that how parents become involved, not simply how much, is what ultimately determines whether involvement benefits a child academically, while Boonk et al. (2018), reviewing 75 studies published between 2003 and 2017, reached a more cautious conclusion overall: not all forms of parental involvement are positively related to achievement, and the relationship is generally less strong and less consistent than is sometimes assumed. Together, these studies caution against treating a positive result for either proxy in the present study as confirmation of parental involvement broadly, since survey response and reported satisfaction capture only a narrow slice of the behaviors these meta-analyses identify as most consequential.

Within the specific context of the xAPI-Edu-Data dataset, the foundational study by Amrieh, Hamtini, and Aljarah (2016) introduced the dataset to the educational data mining community and demonstrated that behavioral features derived from a learning management system — such as the number of times a student raised their hand, visited course resources, or was absent — could meaningfully predict final academic performance when combined with ensemble machine-learning

classifiers. Their study, however, was oriented toward predictive accuracy rather than explanatory inference, and did not isolate the specific contribution of the two parental-involvement variables included in the dataset. This leaves an open, modest gap that the present study is positioned to address: whether these two specific proxies, examined on their own through simple descriptive and inferential statistics, are associated with performance in a way that is consistent with the broader parental-involvement literature summarized above.

The xAPI-Edu-Data set itself grew out of an earlier, smaller study by the same research team.

Amrieh, Hamtini, and Aljarah (2015) first introduced the category of behavioral features — engagement indicators captured automatically by the Kalboard 360 learning-management system, including the two parental-involvement variables used here — in a pilot analysis of 150 students, reporting that classifiers trained with these behavioral features achieved up to a 29% improvement in prediction accuracy compared to classifiers trained without them. The 480-student dataset used in the present study is the expanded version of that same pilot, introduced the following year in Amrieh et al. (2016). More broadly, the use of learning-management-system data of this kind for research on family engagement fits within the wider field of educational data mining, whose methods and typical applications were surveyed by Romero and Ventura (2010); the present study adapts a data-mining-derived dataset to a more traditional explanatory, hypothesis-testing framework rather than the predictive-modeling framework that Romero and Ventura describe as most common in the field.

Taken together, the literature supports two general expectations that motivate the hypotheses of this study. First, given the consistent, if variable, association reported across multiple large-scale meta-analyses (Fan & Chen, 2001; Jeynes, 2007), it is reasonable to expect that at least one of the two involvement proxies available in this dataset will show a statistically significant association with student performance level. Second, because both proxies reflect communicative rather than instrumental involvement (Epstein, 2001), any observed association is more plausibly interpreted as moderate rather than strong, and should be discussed with appropriate caution regarding effect size rather than statistical significance alone.

This expectation is reinforced by two influential research syntheses aimed explicitly at informing school and family policy rather than academic theory alone. Henderson and Mapp (2002), synthesizing 51 studies conducted between 1995 and 2002, concluded that the positive relationship between family involvement and student achievement holds across families of differing race, ethnicity, class, and parental education level, provided that school-led involvement activities are clearly linked to student learning goals.

Desforges and Abouchaar (2003), reviewing the evidence for the UK Department for Education and Skills, reached an even stronger conclusion for younger students: differences in the level of parental involvement in the primary age range were associated with larger differences in achievement than differences in school quality itself, a pattern reported to hold across social classes and ethnic groups. Both syntheses target policy audiences similar to the teachers and school administrators identified as key stakeholders for the present study, and both support treating even indirect, low-resolution proxies of involvement, such as the ones available in xAPI-Edu-Data, as worth investigating rather than dismissing outright.

## 3. Methodology

This study adopts a quantitative, non-experimental, cross-sectional, descriptive-correlational design. The design was selected specifically because it is appropriate for secondary analysis of an existing, already-collected dataset, requires no direct contact with human participants, and can be executed by a single researcher using free, open-source statistical software, without the need for external funding, a research team, or specialized computing infrastructure. No experimental manipulation is involved; the study describes and tests associations between variables as they naturally occur within the dataset.

This design was applied to a single, pre-existing dataset selected specifically for its suitability to secondary, non-experimental analysis. The dataset used in this study is the xAPI-Edu-Data set (also referred to as the Students' Academic Performance Dataset), originally compiled and published by Amrieh, Hamtini, and Aljarah (2016) and made freely available on Kaggle (https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data). The dataset contains 480 student records collected through the Kalboard 360 learning management system, an e-learning platform used in a Jordanian educational context. Each record corresponds to one student and includes 16 variables spanning demographic information (such as gender and nationality), academic background (grade level, section, and course subject), behavioral engagement features derived automatically from the learning management system, and two variables directly describing parental engagement with the school.

Because the dataset is small, anonymized, and already public, its use here involves no additional cost, no new data collection from minors, and no requirement for new institutional ethical approval beyond what was already secured by the dataset's original authors.

Having identified the dataset, attention turns to the specific variables drawn from it and the role each plays in testing the two hypotheses. Table 1 summarizes the variables used in the present analysis. Two variables serve as independent (predictor) variables, operationalized as proxies for parental involvement; three variables serve as control variables, included to characterize the sample and to allow for exploratory secondary comparisons; and one variable, the student's final performance category, serves as the dependent (outcome) variable.

| Variable | Description | Role in the study |
|---|---|---|
| Parent Answering Survey | Whether the parent answered the school's electronic survey (Yes / No) | Independent variable (proxy for parental involvement) |
| Parent school Satisfaction | Parent's reported satisfaction with the school (Good / Bad) | Independent variable (proxy for parental involvement) |
| Raised hands | Number of times the student raised their hand during class sessions | Control variable |
| Visited Resources | Number of times the student accessed course learning resources | Control variable |
| Student Absence Days | Number of school absence days, grouped as Under-7 or Above-7 | Control variable |
| Class | Final academic performance level (Low / Middle / High) | Dependent variable |

With the variables defined, the analytic procedure itself — from data preparation through to hypothesis testing — is outlined in the five steps below. Data were imported into Python (pandas and SciPy) directly from the publicly downloadable CSV file, requiring no preprocessing beyond standard data-cleaning checks (verifying the absence of missing values, confirming correct category labels, and recoding variables into analysis-ready factors where necessary).

Step 1 — Descriptive statistics. Frequencies and percentages were calculated for both parental-involvement proxies (ParentAnsweringSurvey and ParentschoolSatisfaction) and for the dependent variable (Class)
in order to characterize the overall sample composition. Of the N = 480 students, n = 270 (56.2%) had parents who answered the survey and n = 292 (60.8%) reported satisfaction with the school.

Step 2 — Cross-tabulation. Two 2 × 3 contingency tables were constructed: one crossing ParentAnsweringSurvey (Yes/No) with Class (Low/Middle/High), and a second crossing ParentschoolSatisfaction (Good/Bad) with Class. These tables display observed frequencies and row percentages, allowing a direct visual comparison of performance-level distribution across involvement groups.

Step 3 — Inferential testing. A chi-square ($\chi 2$) test of independence was conducted on each of the two contingency tables described above, to formally test hypotheses H1a and H1b against their respective null hypotheses. A conventional significance threshold of $\alpha$=.05 was used. The chi-square statistic is computed as shown in Equation 1, where Oij and Eij denote the observed and expected frequencies in cell i,j:

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} (1)$$

Where the overall chi-square test was statistically significant, Cramér's V was calculated as a standardized measure of effect size, using Equation 2, where N is the total sample size and minr−1, c−1 is the smaller of (rows − 1) and (columns − 1) for the contingency table:

$$V = \sqrt{\frac{\chi^2}{N \times \min(r-1,\, c-1)}} (2)$$

Cramér's V was interpreted using Cohen's (1988) conventional benchmarks for effect sizes based on phi/Cramér's V (small ≈ .10, medium ≈ .30, large ≈ .50 for a 2 × 3 table). Fisher's exact test was planned as a robustness check for any contingency table with expected cell counts below 5 in more than 20% of cells; this threshold was not reached for either table, so the chi-square results are reported directly, without a Fisher's exact supplement.

Step 4 — Exploratory secondary comparisons. As a supplementary step, four continuous engagement indicators (raisedhands, VisitedResources, AnnouncementsView, and Discussion) were compared across involvement groups using Mann-Whitney U tests, given the non-normal distribution of these variables; this comparison was run separately for each proxy of parental involvement (ParentAnsweringSurvey and ParentschoolSatisfaction), in order to explore whether more

involved parents are also associated with different levels of observable student engagement. StudentAbsenceDays, being categorical (Under-7/Above-7), was instead compared descriptively as the percentage of students exceeding seven absence days in each involvement group. Two further exploratory analyses probed the robustness and interpretation of these findings: a robustness check repeating the H1a comparison separately for students whose survey response came from the father versus the mother (Relation), and a joint-proxy analysis combining both involvement indicators into a single three-category variable (positive on both, negative on both, or mixed) to test whether the two proxies capture a shared underlying pattern of family engagement.

Step 5 — Visualization. Simple, easily interpretable bar charts and boxplots were produced to visually represent the distribution of performance levels and engagement indicators across involvement groups, for inclusion in the Results section and for use in any non-technical presentation of the findings to school stakeholders.

Before turning to the Results, one further methodological point deserves attention: the ethical status of reusing an existing, third-party dataset in this way. Because this study relies exclusively on a pre-existing, publicly available, fully anonymized secondary dataset, no new data were collected from students, parents, or school staff, and no personally identifiable information is contained in or derived from the dataset. The original data collection was conducted by the dataset's authors under the ethical and institutional procedures applicable to the source learning management system; the present study does not require separate participant consent or additional institutional review, consistent with standard practice for secondary analysis of open, de-identified educational datasets. Researchers reusing this dataset should nonetheless cite the original source (Amrieh, Hamtini, & Aljarah, 2016) and avoid any attempt to re-identify individual students.

It is worth noting from the outset — and returning to in the Discussion and Limitations sections — that both parental-involvement variables in this dataset are binary and self-reported at a single point in time, which limits the depth and precision with which "involvement".

## 4. Results

This section presents the results of the analysis described in the Methodology, based on the publicly available xAPI-Edu-Data dataset (N = 480). Both hypotheses were tested using chi-square tests of independence, and exploratory comparisons on additional engagement variables were conducted using Mann-Whitney U tests.

The presentation begins with a descriptive overview of the sample before turning to the formal tests of H1a and H1b. Of the N = 480 students in the sample, 270 (56.2%) had parents who answered the school survey, while 210 (43.8%) did not. Regarding satisfaction, 292 (60.8%) had parents who reported being satisfied with the school ("Good"), while 188 (39.2%) reported dissatisfaction ("Bad"). Regarding academic performance, 127 (26.5%) of students were classified as Low, 211 (44.0%) as Middle, and 142 (29.6%) as High. Table 2 presents these descriptive statistics in full.

**Table 2. Descriptive Statistics (N = 480)**

| Variable | n | % |
|---|---|---|
| Parent Answered Survey (Yes) | 270 | 56.2% |
| Parent Answered Survey (No) | 210 | 43.8% |
| Parent School Satisfaction (Good) | 292 | 60.8% |
| Parent School Satisfaction (Bad) | 188 | 39.2% |

With the sample profile established, the first hypothesis — concerning parental survey response — can now be tested directly. Table 3 presents the 2×3 contingency table crossing ParentAnsweringSurvey with Class. Among students whose parents did not answer the survey, 99 (47.1%) were classified as Low, 83 (39.5%) as Middle, and only 28 (13.3%) as High. Among students whose parents did answer the survey, the pattern reversed: only 28 (10.4%) were Low, 128 (47.4%) were Middle, and 114 (42.2%) were High. A chi-square test of independence indicated that this relationship was statistically significant, $\chi 22, N=480=95.36, p<.001$, Cramér's V=.45, representing a large effect. Hypothesis H1a was therefore supported.

**Table 3. Parental Survey Response × Academic Performance Level (N = 480)**

| ParentAnsweringSurvey | Low | Middle | High |
|---|---|---|---|
| No (n = 210) | 99 (47.1%) | 83 (39.5%) | 28 (13.3%) |
| Yes (n = 270) | 28 | 128 | 114 |

| | (10.4%) | (47.4%) | (42.2%) |
|---|---|---|---|
| $\chi^2(2, N = 480) = 95.36, p < .001, V = .45$ | | | |

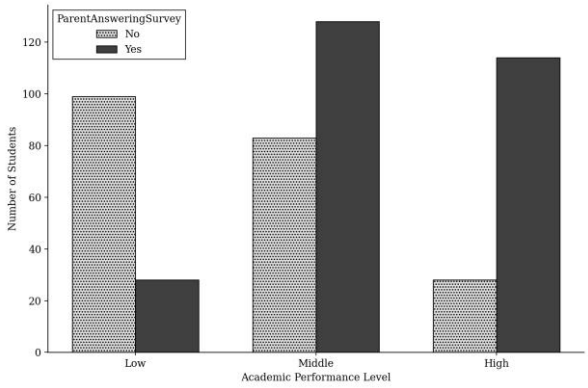


**Figure 1. Parent Survey Response by Academic Performance Level**

The second parental-involvement proxy, school satisfaction, was tested in the same way. Table 4 presents the 2×3 contingency table crossing ParentschoolSatisfaction with Class. Among students whose parents reported dissatisfaction ("Bad"), 84 (44.7%) were classified as Low, 80 (42.6%) as Middle, and 24 (12.8%) as High. Among students whose parents reported satisfaction ("Good"), only 43 (14.7%) were Low, 131 (44.9%) were Middle, and 118 (40.4%) were High. A chi-square test of independence indicated that this relationship was also statistically significant, $\chi 22, N=480=68.47, p<.001$, Cramér's V=.38, representing a medium-to-large effect. Hypothesis H1b was therefore also supported.

**Table 4. Parental School Satisfaction × Academic Performance Level (N = 480)**

| ParentschoolSatisfaction | Low | Middle | High |
|---|---|---|---|
| Bad (n = 188) | 84 (44.7%) | 80 (42.6%) | 24 (12.8%) |
| Good (n = 292) | 43 (14.7%) | 131 (44.9%) | 118 (40.4%) |
| $\chi^2(2, N = 480) = 68.47, p < .001, V = .38$ | | | |

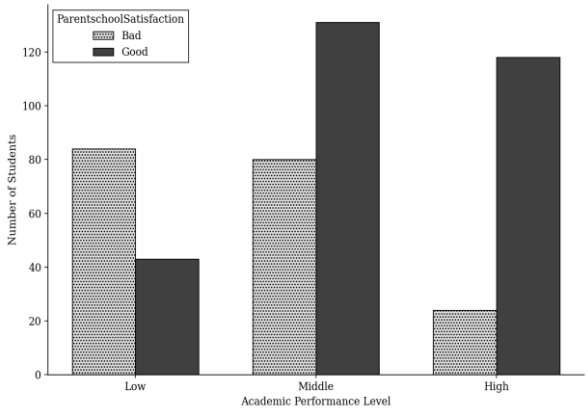


**Figure 2. Parent School Satisfaction by Academic Performance Level**

To further contextualize these findings, exploratory Mann-Whitney U tests compared classroom engagement indicators between students whose parents did and did not answer the survey (Table 5). Students with survey-responding parents raised their hand in class more often (Mdn=69.5 vs. 25.0, U=38718.5, p<.001), visited more learning resources (Mdn=79.0 vs. 29.5, U=40470.0, p<.001), viewed more course announcements (Mdn=49.0 vs. 20.0, U=41641.5, p<.001), and participated more in class discussion (Mdn=48.5 vs. 30.0, U=35852.0, p<.001). Students with survey-responding parents were also considerably less likely to have more than seven days of absence (77 of 270, 28.5%) compared with students whose parents did not answer the survey (114 of 210, 54.3%).

**Table 5. Exploratory Comparisons of Engagement Indicators by Parental Survey Response**

| Variable | Mdn (No) | Mdn (Yes) | Mann-Whitney U, p |
|---|---|---|---|
| raisedhands | 25.0 | 69.5 | $U = 38718.5, p < .001$ |
| VisitedResources | 29.5 | 79.0 | $U = 40470.0, p < .001$ |
| AnnouncementsView | 20.0 | 49.0 | $U = 41641.5, p < .001$ |
| Discussion | 30.0 | 48.5 | $U = 35852.0, p < .001$ |

The engagement comparisons above focused on ParentAnsweringSurvey. To test whether ParentschoolSatisfaction shows a similar behavioral signature, the same Mann-Whitney U comparisons were repeated using satisfaction group as the grouping variable (Table 6). Students whose parents reported satisfaction with the school raised their hand more often (Mdn=69.0 vs. 27.0, U=37,142.0, p<.001), visited more learning resources (Mdn=80.0 vs. 33.0, U=39,182.5, p<.001), and viewed more course announcements (Mdn=42.0 vs. 20.5, U=37,464.5, p<.001) than students whose parents reported dissatisfaction. Participation in class discussion, however, did not differ significantly between the two satisfaction groups (Mdn=40.0 vs. 35.0, U=29,220.5, p=.232) — in contrast to the survey-response comparison above, where discussion participation did differ significantly (Figure 5). Absence patterns followed the same direction as for survey response: students with dissatisfied parents were more likely to have more than seven days of absence (101 of 188, 53.7%) than students with satisfied parents (90 of 292, 30.8%).

**Table 6. Engagement Differences by Parental School Satisfaction**

| Variable | Mdn (Bad) | Mdn (Good) | Mann-Whitney U, p |
|---|---|---|---|
| Raised hands | 27.0 | 69.0 | $U = 37142.0, p < .001$ |
| Visited Resources | 33.0 | 80.0 | $U = 39182.5, p < .001$ |
| AnnouncementsView | 20.5 | 42.0 | $U = 37464.5, p < .001$ |
| Discussion | 35.0 | 40.0 | $U = 29220.5, p = .232$ |

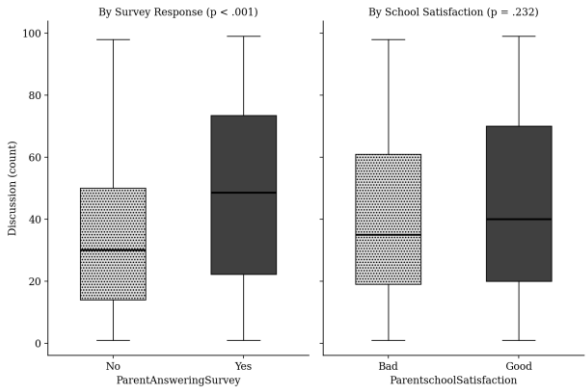


**Figure 5. Class Discussion Participation by Involvement Proxy**

A further question is whether the H1a association holds regardless of which parent — mother or father — responded to the survey. The Relation variable in xAPI-Edu-Data records whether the parent who completed the school survey was the student's mother or father. As a partial check on whether the H1a association is driven by one parent group, the chi-square

test for ParentAnsweringSurvey and Class was re-run separately within the father subsample (n = 283) and the mother subsample (n = 197) (Table 7). The association remained statistically significant and of comparable magnitude in both groups: $\chi 22,n=283=55.42,p<.001$, Cramér's V=.44 when the father was the responding parent, and $\chi 22,n=197=30.67,p<.001$, Cramér's V=.40 when the mother was the responding parent. This pattern is consistent with Kim and Hill's (2015) meta-analytic finding that the strength of the involvement–achievement relationship does not differ meaningfully by parent gender, and suggests that the H1a result reported above is not an artifact of either parent subgroup alone. This check does not, however, resolve whether it is specifically the responding parent's own involvement that matters, since the dataset does not indicate whether the non-responding parent in each household was also uninvolved.

**Table 7. Parental Survey Response × Academic Performance Level, by Responding Parent (Relation)**

| Subgroup | Low | Middle | High |
|---|---|---|---|
| Father, No (n = 143) | 80 (55.9%) | 57 (39.9%) | 6 (4.2%) |
| Father, Yes (n = 140) | 24 (17.1%) | 80 (57.1%) | 36 (25.7%) |
| Mother, No (n = 67) | 19 (28.4%) | 26 (38.8%) | 22 (32.8%) |
| Mother, Yes (n = 130) | 4 (3.1%) | 48 (36.9%) | 78 (60.0%) |

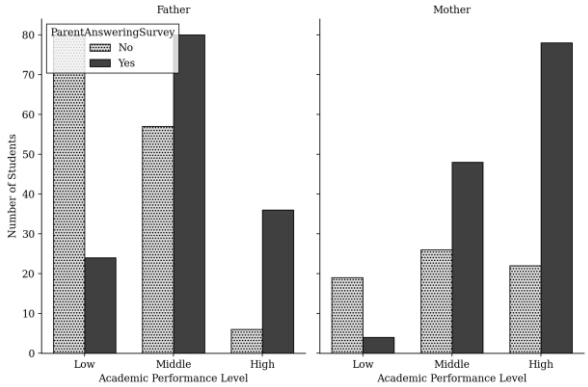


**Figure 3. Parental Survey Response by Academic Performance Level, by Responding Parent (Relation)**

Because ParentAnsweringSurvey and ParentschoolSatisfaction are conceptually related but not identical — a parent can respond to a school survey without reporting satisfaction, or vice versa — a supplementary analysis examined their joint distribution. Students were grouped into three categories: positive on both proxies (n = 227), negative on both (n = 145), and mixed, positive on exactly one proxy (n = 108) (Table 8). The association between this three-level joint indicator and academic performance level was itself statistically significant, $\chi 24,N=480=109.41,p<.001$, Cramér's V=.34. The gap between the two consistent groups was substantial: 43.6% of students in the both-positive group were classified as High performers, compared with only 6.2% in the both-negative group; conversely, 54.5% of the both-negative group were classified as Low performers, compared with 10.1% of the both-positive group. Students in the mixed group fell consistently between these two extremes (31.5% High, 23.1% Low), suggesting that the two proxies capture at least partially independent information rather than one simply standing in for the other, and that their joint presence identifies a more sharply delineated group at either end of the performance distribution than either proxy alone.

**Table 8. Joint Involvement Proxy × Academic Performance Level (N = 480)**

| Joint Group | Low | Middle | High |
|---|---|---|---|
| Both negative (n = 145) | 79 (54.5%) | 57 (39.3%) | 9 (6.2%) |
| Mixed (n = 108) | 25 (23.1%) | 49 (45.4%) | 34 (31.5%) |
| Both positive (n = 227) | 23 (10.1%) | 105 (46.3%) | 99 (43.6%) |

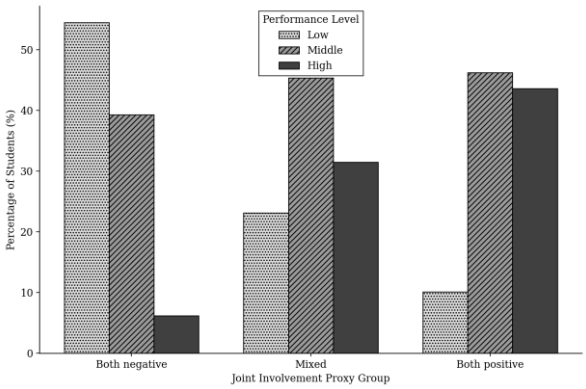


**Figure 4. Joint Involvement Proxy by Academic Performance Level**

In summary, both hypotheses were supported: students whose parents answered the school survey or reported satisfaction with the school were significantly more likely to fall into higher academic performance categories, with effect sizes ranging from medium-to-large (H1b) to large (H1a). These associations were accompanied by parallel differences in observable classroom engagement and attendance, which are explored further in the Discussion below. The supplementary analyses further showed that this engagement signature held for both proxies individually — with the exception of class discussion participation, which distinguished survey-response groups but not satisfaction groups — that the association was robust across both mother- and father-responding subsamples, and that combining the two proxies identified students at the extremes of the performance distribution with sharper precision than either proxy alone.

**5. Discussion**

The present analysis found that both parental-involvement proxies available in the xAPI-Edu-Data dataset — whether a parent answered the school survey and whether a parent reported satisfaction with the school — were significantly and substantially associated with students' academic performance level, with the survey-response proxy showing a large effect (V=.45) and the satisfaction proxy a medium-to-large effect (V=.38).

These results are broadly consistent with the meta-analytic evidence reported by Fan and Chen (2001) and Jeynes (2007), and support Epstein's (2001) framing of parent-school communication as a meaningful dimension of family engagement, even when measured through simple, low-resolution binary indicators rather than detailed validated scales. Notably, the observed effect sizes in this study (V=.45 and V=.38) are considerably larger than the small-to-moderate correlations typically reported in the broader literature reviewed above; this gap is discussed further below.

For context, the most comprehensive recent synthesis of this literature, Barger, Kim, Kuncel, and Pomerantz's (2019) meta-analysis of 448 independent studies encompassing 480,830 families, reported small positive associations between parents' naturally occurring involvement and children's academic adjustment, with correlations ranging from r=.13 to r=.23.

Any Cramér's V obtained in the present analysis should be interpreted against this modest benchmark rather than against the larger effect sizes sometimes reported for more intensive, expectation-based forms of involvement (Wilder, 2014; Boonk et al., 2018).

Even so, the fact that two such indirect and coarse proxies — a single yes/no survey response and a single-item satisfaction rating — produced effects this large is notable, since prior research using more detailed, validated involvement scales has generally found smaller associations (Fan & Chen, 2001; Barger et al., 2019). One plausible explanation is that ParentAnsweringSurvey and ParentschoolSatisfaction are not measuring involvement in isolation, but are acting as broader proxies for a cluster of supportive home behaviors and parent-school communication patterns — the very mechanisms Epstein (2001) and Henderson and Mapp (2002) describe as central to family engagement.

This clustering interpretation receives some direct support from the joint-proxy analysis reported above (Table 8). If ParentAnsweringSurvey and ParentschoolSatisfaction were simply two noisy measurements of the same underlying involvement, their combination would be expected to add little beyond what either proxy shows alone. Instead, students who were positive on both proxies and students who were negative on both proxies diverged more sharply (43.6% vs. 6.2% classified as High performers) than the gap observed for either individual proxy, while the mixed group — positive on only one — fell in between rather than resembling either extreme. This pattern is more consistent with the two proxies indexing partially distinct components of family engagement that compound when both are present, than with a single latent involvement factor measured twice.

This interpretation is consistent with Patall, Cooper, and Robinson's (2008) research synthesis on parental involvement in homework specifically, which found that active, communication-based forms of involvement tend to matter more than passive monitoring. It is also consistent with Castro et al. (2015), who found that the strongest associations with achievement emerge specifically when families communicate about school activities and hold high expectations, rather than merely responding to a survey or reporting general satisfaction — which is precisely the kind of active engagement that a parent who both answers a school survey and reports being satisfied is likely to exhibit.

The exploratory comparisons on control variables reported above (Table 5) further help contextualize the main findings. Students whose parents answered the survey also raised their hand more often, visited more learning resources, viewed more course announcements, participated more in discussion, and were absent less frequently. This pattern suggests a possible behavioral mechanism linking parental engagement to performance: rather than directly shaping test scores, involved parents may indirectly shape their children's day-to-day classroom engagement, which in turn translates into stronger

academic outcomes — a pathway consistent with Pomerantz, Moorman, and Litwack's (2007) account of parental involvement operating through children's motivation and self-regulation rather than through direct instruction.

One nuance in this behavioral pattern deserves note. Class discussion participation was significantly higher among students whose parents answered the school survey (Table 5), but not among students whose parents reported satisfaction (Table 6). This divergence suggests that the two involvement proxies, despite each being associated with better classroom engagement overall, are not fully interchangeable. Answering a school survey may mark a more actively communicative parenting style — one that also correlates with a child's willingness to speak up in class — whereas reporting satisfaction may capture a more passive, retrospective evaluation of the school that is less tied to encouraging classroom voice specifically. If so, researchers and school staff should be cautious about treating these two commonly available proxies as fully equivalent indicators of parental involvement, even though both were significantly associated with academic performance in this study.

From a policy standpoint, this potential mechanism would align with Henderson and Mapp's (2002) recommendation that schools link parent-involvement activities directly to student learning goals rather than treating involvement as valuable in isolation, and with Desforges and Abouchaar's (2003) finding that, at least for younger students, differences in parental involvement can matter as much as, or more than, differences in school quality.

From a practical standpoint, and regardless of the specific direction of the findings, this study illustrates that even a single researcher, working without external funding, specialized software, or institutional data-collection infrastructure, can produce a small, transparent, and replicable contribution to the applied literature on family engagement by making use of open educational datasets. Given that a simple, low-cost measure such as parental survey response was associated with a large difference in academic outcomes, schools may benefit from treating survey non-response itself as an early, low-cost signal for outreach, prioritizing additional communication with families who do not typically respond to routine school surveys.

**6. Limitations**

This study has several limitations that should be considered when interpreting its findings. First, the design is cross-sectional and correlational; even where statistically significant associations are found, no causal claims can be made about the direction of the relationship between parental involvement and academic performance. It is equally plausible, for instance, that parents become more engaged with the school in response to a child's strong or weak performance, rather than involvement causing performance.

Second, both parental-involvement variables available in the dataset are binary, self-reported, and captured at a single point in time, offering only a coarse approximation of a construct that the literature treats as multidimensional and context-dependent (Epstein, 2001). Richer, validated survey instruments capturing multiple dimensions of involvement were not available in this secondary dataset.

A related limitation concerns which parent's response the dataset records. The Relation variable in xAPI-Edu-Data indicates whether the survey respondent was the student's mother or father; a robustness check reported above (Table 7) found that the H1a association held with comparable magnitude in both the father-responding and mother-responding subsamples, consistent with Kim and Hill's (2015) meta-analytic finding that the involvement-achievement relationship is statistically similar for fathers and mothers despite mothers reporting higher average involvement; however, this check cannot determine whether it is specifically the responding parent's own involvement that matters. The dataset also offers no measure of homework-specific involvement, one of the most extensively studied forms of parental involvement in the literature (Patall et al., 2008), nor does it include students young enough to speak to the early-childhood and early-elementary populations examined by Ma, Shen, Krenn, Hu, and Yuan (2016), whose meta-analysis found that the strength of the parental-involvement–achievement relationship can vary by developmental stage. Consequently, the present findings should be understood as applying specifically to the two communicative proxies available for this age range, not as a general test of parental involvement as a construct.

Third, the sample (N = 480) originates from a single learning management system used in a specific Jordanian educational context (Amrieh, Hamtini, & Aljarah, 2016), which may limit the generalizability of the findings to other educational systems, cultural contexts, or age groups. Replication with independent samples from different regions and educational levels would be needed before drawing broader conclusions.

Finally, the dataset does not include information on socioeconomic status, parental education level, or family structure, all of which are known confounding variables in the parental-involvement literature (Fan & Chen, 2001) and could not be statistically controlled for in the present analysis.

This omission is consequential because Lee and Bowen (2006), drawing on Bourdieu's theory of cultural capital, argued that parents with greater cultural capital tend to be both more involved in their children's schooling and more effective when they are involved, such that failing to account for cultural capital and socioeconomic status can distort the apparent relationship between involvement indicators, such as the ones used here, and academic achievement. Because xAPI-Edu-Data provides no socioeconomic or cultural-capital measure, the present study cannot rule out that any observed association between ParentAnsweringSurvey or ParentschoolSatisfaction and academic performance is partly confounded by unmeasured family background factors of this kind.

**7. Conclusion**

This study examined whether two simple, low-resolution proxies of parental involvement — survey response and reported school satisfaction — were associated with academic performance among 480 secondary school students. Both proxies were significantly and substantially associated with performance level, with students of involved parents far more likely to be classified in the Middle or High performance categories and to show higher classroom engagement and lower absenteeism.

For schools, a practical, low-cost implication is that tracking parental survey non-response can serve as an early flag for targeted family outreach. Future research should replicate this analysis on more recent or geographically distinct datasets, and should incorporate richer, validated measures of parental involvement — such as academic expectations or direct homework support — to determine whether they explain an even larger share of the variance in student performance than the coarse proxies available here.

**8. Declarations**

Funding: This research received no external funding. It was conducted independently, without institutional or grant support.
Conflict of Interest: The author declares no conflict of interest.
Data Availability: The dataset analyzed in this study (xAPI-Edu-Data) is publicly available on Kaggle at https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data and was originally compiled by Amrieh, Hamtini, and Aljarah (2016). No new data were collected.

**References**

1. Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November). Preprocessing and analyzing educational data set using X-API for improving student's performance. In 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1–5). IEEE. https://doi.org/10.1109/AEECT.2015.7360581
2. Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. International Journal of Database Theory and Application, 9(8), 119–136.
3. Barger, M. M., Kim, E. M., Kuncel, N. R., & Pomerantz, E. M. (2019). The relation between parents' involvement in children's schooling and children's adjustment: A meta-analysis. Psychological Bulletin, 145(9), 855–890. https://doi.org/10.1037/bul0000201
4. Boonk, L., Gijselaers, H. J. M., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. Educational Research Review, 24, 10–30. https://doi.org/10.1016/j.edurev.2018.02.001
5. Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-analysis. Educational Research Review, 14, 33–46. https://doi.org/10.1016/j.edurev.2015.01.002
6. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
7. Desforges, C., & Abouchaar, A. (2003). The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review (Research Report No. 433). Department for Education and Skills.
8. Epstein, J. L. (2001). School, family, and community partnerships: Preparing educators and improving schools. Westview Press.
9. Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. Educational Psychology Review, 13(1), 1–22. https://doi.org/10.1023/A:1009048817385
10. Henderson, A. T., & Mapp, K. L. (2002). A new wave of evidence: The impact of school, family, and community connections on student achievement. National Center for Family and Community Connections with Schools, Southwest Educational Development Laboratory.
11. Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. Developmental Psychology, 45(3), 740–763. https://doi.org/10.1037/a0015362
12. Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. Urban Education, 40(3), 237–269. https://doi.org/10.1177/0042085905274540
13. Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. Urban Education, 42(1), 82–110. https://doi.org/10.1177/0042085906293818
14. Kim, S., & Hill, N. E. (2015). Including fathers in the picture: A meta-analysis of parental involvement and students' academic achievement. Journal of Educational Psychology, 107(4), 919–934. https://doi.org/10.1037/edu0000023
15. Lee, J.-S., & Bowen, N. K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. American Educational Research Journal, 43(2), 193–218. https://doi.org/10.3102/00028312043002193
16. Ma, X., Shen, J., Krenn, H. Y., Hu, S., & Yuan, J. (2016). A meta-analysis of the relationship between learning outcomes and parental involvement during early childhood education and early elementary education. Educational Psychology Review, 28(4), 771–801. https://doi.org/10.1007/s10648-015-9351-1
17. Patall, E. A., Cooper, H., & Robinson, J. C. (2008). Parent involvement in homework: A research synthesis. Review of Educational Research, 78(4), 1039–1101. https://doi.org/10.3102/0034654308325185
18. Pomerantz, E. M., Moorman, E. A., & Litwack, S. D. (2007). The how, whom, and why of parents' involvement in children's academic lives: More is not always better. Review of Educational Research, 77(3), 373–410. https://doi.org/10.3102/003465430305567
19. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532
20. Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. Educational Review, 66(3), 377–397. https://doi.org/10.1080/00131911.2013.780009