Exploring the Relationship between Library Utilization and Academic Success Using a Composite Library Use Index

2026 JULAC LIBRARIES FORUM

9 JAN 2026

FRANCIS CHAN

THE EDUCATION UNIVERSITY OF HONG KONG LIBRARY

Q&A

Research Question

Q&A

What do we want to find out?

- Is there any correlation between students' library usages and their academic success?

We are looking for objective data and evidence to demonstrate the library's impact and value in supporting student success.

Q&A

Background: Past Studies

Q&A

Past Studies

Many studies have been conducted before, dating back to the 1960's.

Most have shown some positive correlation between library utilization and student GPA or lower drop out rate.

Borrowing statistics and access logs to online resources are the most often used indicators for library usage. Library entrance figures, workshop attendance, hours logged in to library computers, etc. have also been used.

In addition to statistical analysis of logs and data, methods such as focus groups and questionnaire surveys have also been used in some studies.

Q&A

Local Studies

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

More than a decade ago, Hong Kong Baptist University Library has conducted two studies:

Wong, S., & Webb, T. (2011). Uncovering Meaningful Correlation between Student Academic Performance and Library Material Usage. *College & Research Libraries, 72*(4), 361-370. doi: https://doi.org/10.5860/crl-129

Wong, S., & Cmor, D. (2011). Measuring Association between Library Instruction and Graduation GPA. *College & Research Libraries, 72*(5), 464-473. doi: https://doi.org/10.5860/crl-151

Major findings:

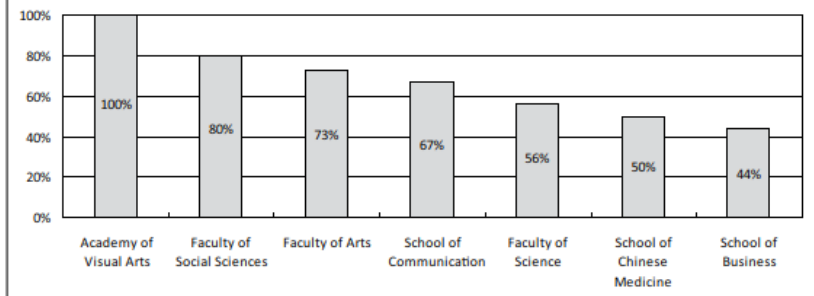
Statistical significant correlations between borrowing data and students' graduation GPA were found in 31 of 48 academic units

Significant correlations between library workshop attendance and students' graduation GPA were found for those academic units with three or more library workshops being offered

Q&A

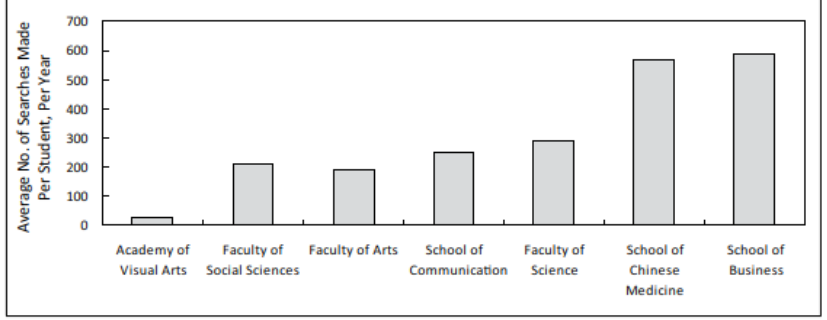
One Interesting Finding from Wong & Webb 2011 Study …

FIGURE 3
Percentage of Sample Groups That Have a Positive Correlation in the Corresponding Population (Faculty/School Based)



Note: Academy of Visual Arts has only one valid sample group.

FIGURE 4
Usage of Electronic Resources (Faculty/School Based)



It's not a coincidence that those Faculties / Schools with less sample groups with positive correlation between student GPA and borrowing statistics also are the ones with the highest average use of electronic resources.

In other words:
If we want to get the whole picture of the relationship between library use and academic performance, we should look at more than one kind of library use.

Q&A

Our Study

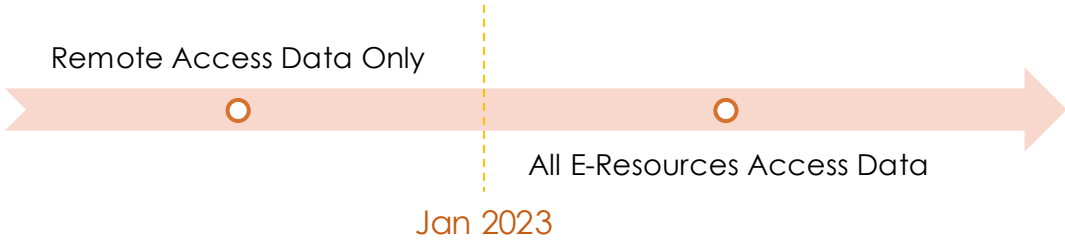
Q&A

The Current EdUHK Library Study – Hypothesis and Timing

- **Research Hypothesis**
  - Higher Library Use ←→ Higher Academic Success?

  (Correlation, not Causal Relationship)

- **Why Now?**
  - Comprehensive e-resources usage data became available when a mandatory login policy for all library e-resources regardless of location was implemented in Jan 2023.



Q&A

The Current EdUHK Library Study – Data and Methodology

- **Data**
  - Student GPA and background data collected from University's academic information system.
  - 5 kinds of library usage data collected by the Library.
- **Methodology**
  - Generalized Linear Mixed Model (GLM), a multivariate statistical model, will be used.

Q&A

Student Data

Q&A

Student Data Collected

Student information for 2024/25 academic year obtained from University's academic information system:

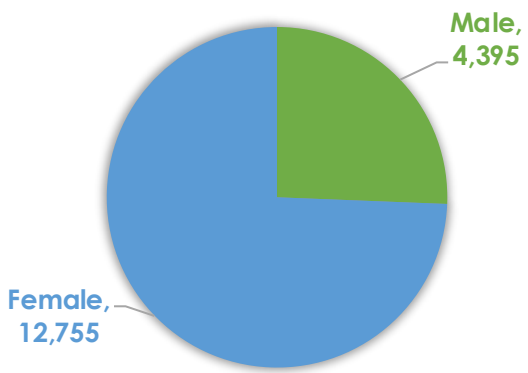
- Masked student ID
- Term GPA*
- Cumulative GPA*
- Programme of Study
- Year of Study
- Gender
- Residency
- Hostel

* ***Term GPA was used*** *instead of Cumulative GPA for this study because our library usage data were also term-based figures.*

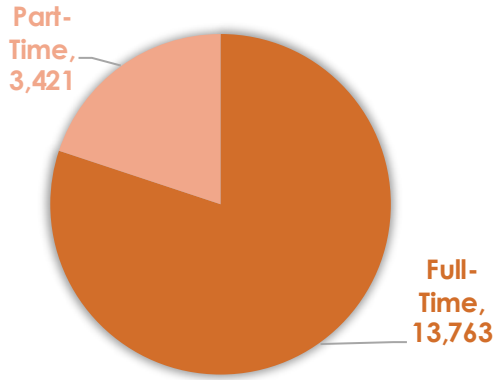
Q&A

Student Data Distribution

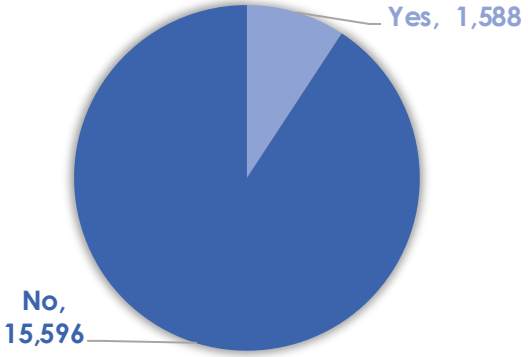
GENDER



MODE OF STUDY



LIVING IN HOSTEL



Total data points with Term GPA*
= 17,184 (1st Term: 8,556 + 2nd Term: 8,628)

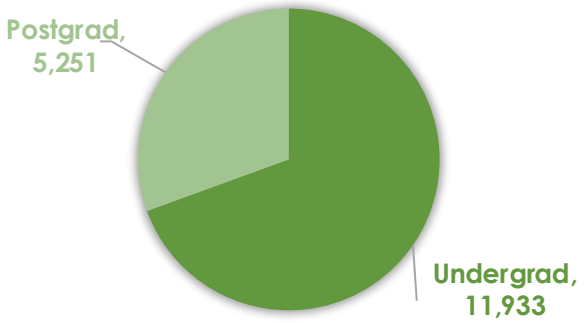
** Some programmes were excluded for this study:*

- *Small programmes with less than 10 students / year of study / term; and*
- *Programmes composed of students from different academic disciplines under the same Programme Code (e.g. Doctor of Philosophy, Master of Philosophy)*

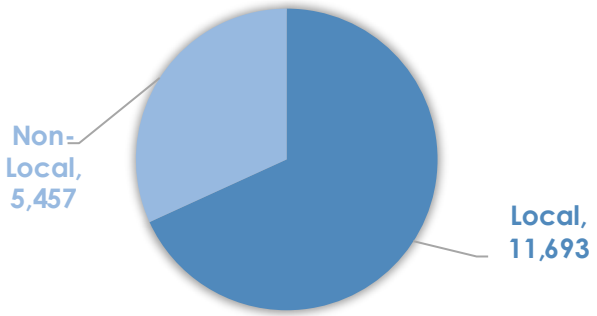
Q&A

Student Data Distribution

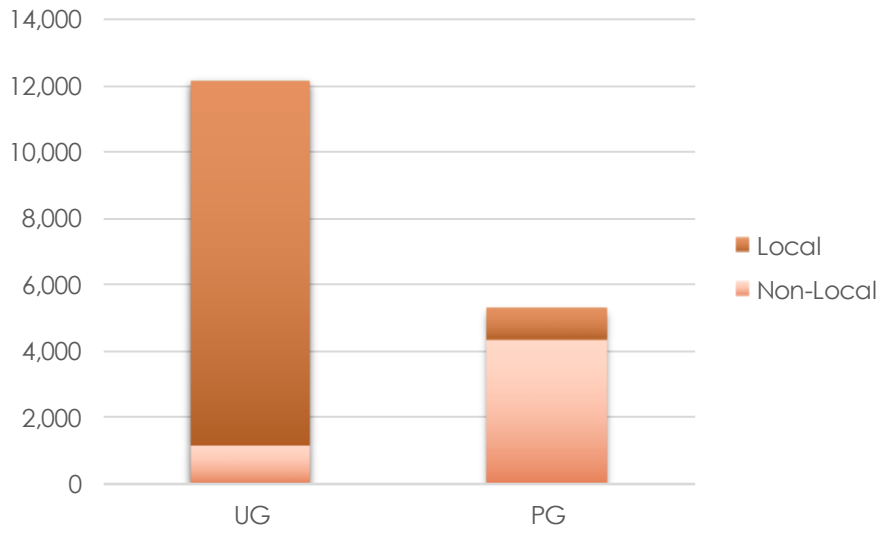
PROGRAMME LEVEL



RESIDENCY



	Local : Non-Local
Undergrad	9 : 1
Postgrad	2 : 8
Overall	7 : 3



Q&A

Library Data

Q&A

Library Data Collected

- 5 Kinds of Library Data were Used in this Study:

Loan (loan counts)

Entrance (library entrances)

E-Sessions (electronic resources sessions)

Booking (library facilities booking)

Workshop (library workshop attendance)

- Student Information and Library Use Data were matched using the masked ID.

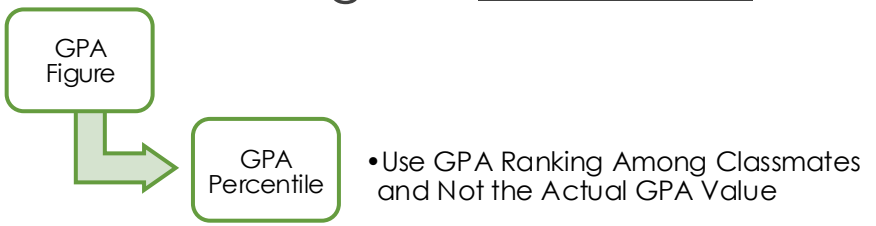
Q&A

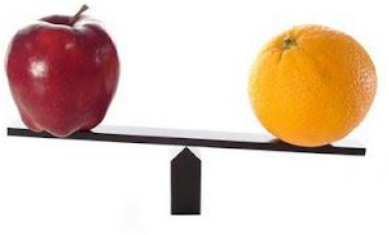
Data Transformation

Q&A

Data Transformation: GPA

- **Transformation of Student GPA**
  - To avoid comparing apples and oranges (different departments may have different grading guidelines and standards), we transformed the GPA number to GPA percentile for the students in the same programme, same year of study during the same term.





Q&A

GPA Normalization Computation Example

Term	Programme	Year of Study	Student	GPA	GPA Percentile
1st Term	X	1	A	3.25	84.85
2nd Term	X	1	A	3.26	61.86
1st Term	X	4	B	3.25	38.10
2nd Term	X	4	B	3.33	63.81
1st Term	Y	1	C	3.25	66.67
2nd Term	Y	1	C	2.79	13.76
1st Term	Y	4	D	3.25	62.04
2nd Term	Y	4	D	3.19	54.39

A GPA of 3.25 may be ranked as high as 85th percentile for one class of students but just 38th percentile in another

Q&A

Data Transformation: Library Use Data

- **Why Transformation Needed?**
  - Students from different disciplines may use the library in different ways. Some students may be heavy users of online journals but never borrow a book.
  - To avoid comparing apples and oranges, we may build 5 statistical models for the 5 kinds of library usage data we collected and then explain why loan statistics works better for Programme A but e-resources data should be used for Programme B.
  - But we chose the alternative by constructing a Composite Library Use Index based on transformed (normalized) library data.

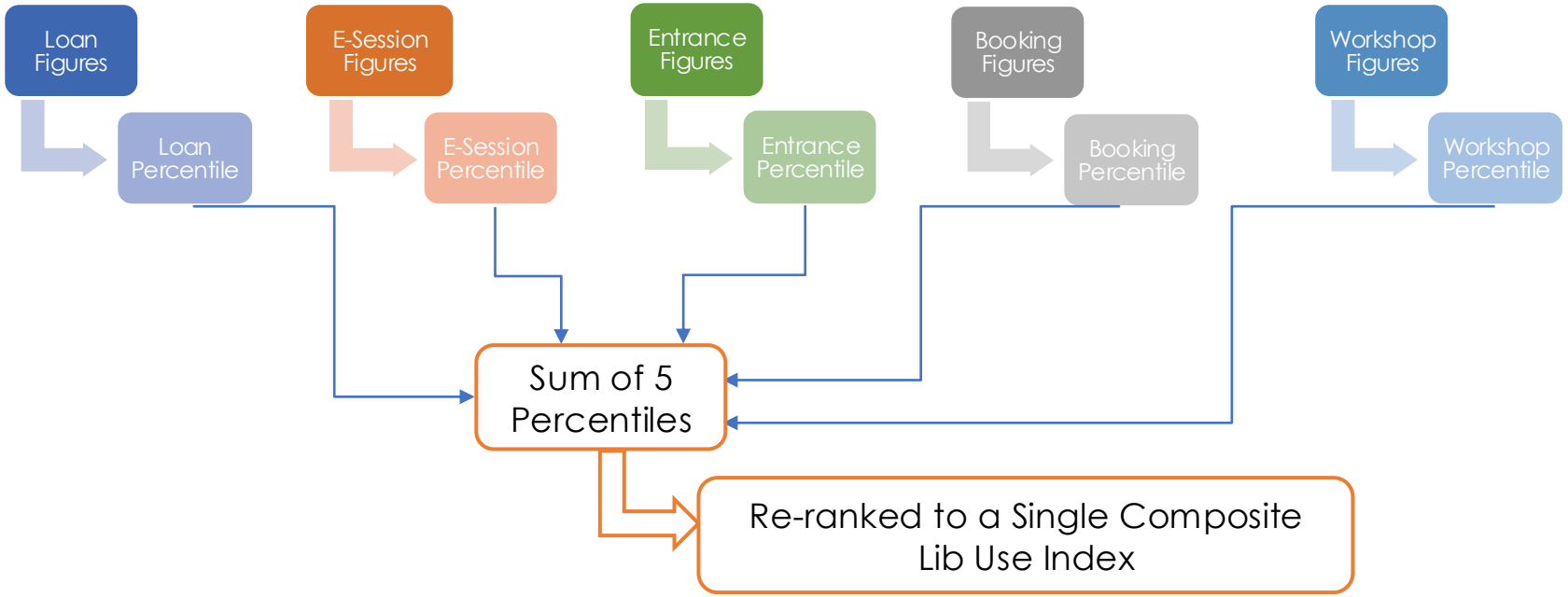
Q&A

Data Transformation: Library Use Data

- **Constructing a Composite Library Use Index**
  - We construct a composite library use index in 3 steps:
    1. Transforming each of the five library use data from the actual figure to a relative ranking percentile (compared to classmates in the same programme, same year of study during the same term).
    2. Add the five library usage percentiles.
    3. Obtain the Composite Library Use Index by finding the relative ranking percentile for the sum from Step 2, compared to their classmates.

Q&A

Transforming 5 Kinds of Library Data into a Single Composite Library Use Index



Q&A

Library Use Normalization Computation Example

Term	Prog	Year of Study	Student	Loan	Entrance	Booking	ESessions	Workshops	Loan (%)	Entrance (%)	Booking (%)	ESessions (%)	Workshops (%)	ALL LIB USE (%)	Composite Lib Use Index
1st Term	X	1	A	3	19	1	6	0	91.74	62.39	65.14	77.98	0.92	298.17	93.58
2nd Term	X	1	A	5	23	0	19	2	66.67	70.18	0.88	89.47	96.49	323.69	89.47
1st Term	X	4	B	38	8	1	6			30.70	76.32	53.51	91.23	347.37	94.74
2nd Term	X	4	B	3	19	0	6			43.86			9		78.95
1st Term	Y	1	C	0	2	0	3			5.06	1.01	73.74	98.99	179.80	53.54
2nd Term	Y	1	C	0	3	0	2			3.09	1.03	58.76	95.88	159.79	50.52
1st Term	Y	4	D	5	16	0	4	4	70.48	52.38	0.95	9.52	92.38	225.71	70.48
2nd Term	Y	4	D	21	15	0	20	4	91.43	43.81	0.95	35.24	92.38	263.81	77.14

3 → 91.74%
19 → 62.39%
1 → 65.14%
6 → 77.98%
0 → 0.92%

298.17% → 93.58

Q&A

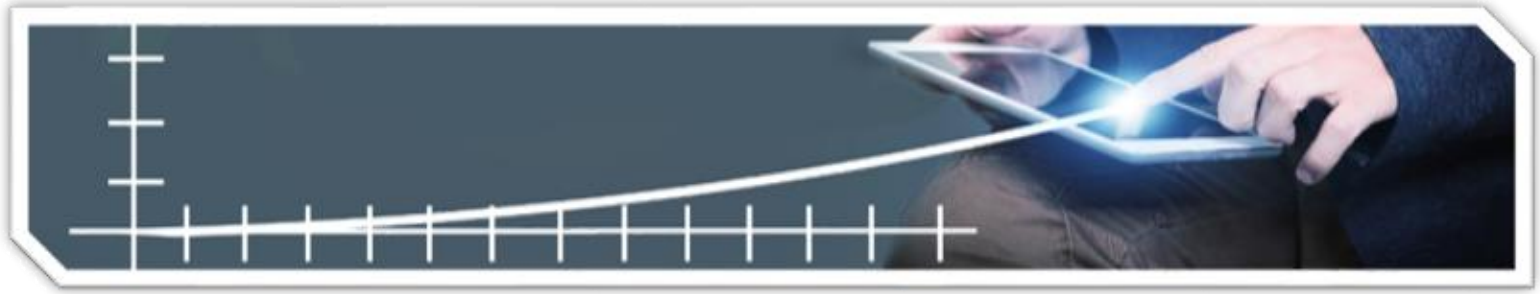
Data Transformation:
GPA + Lib Use Data

- In Short, we are exploring the correlation of the students' relative level of combined library usage and the relative academic performance as compared to their classmates during the same school term.

Absolute Value → Percentile Rank

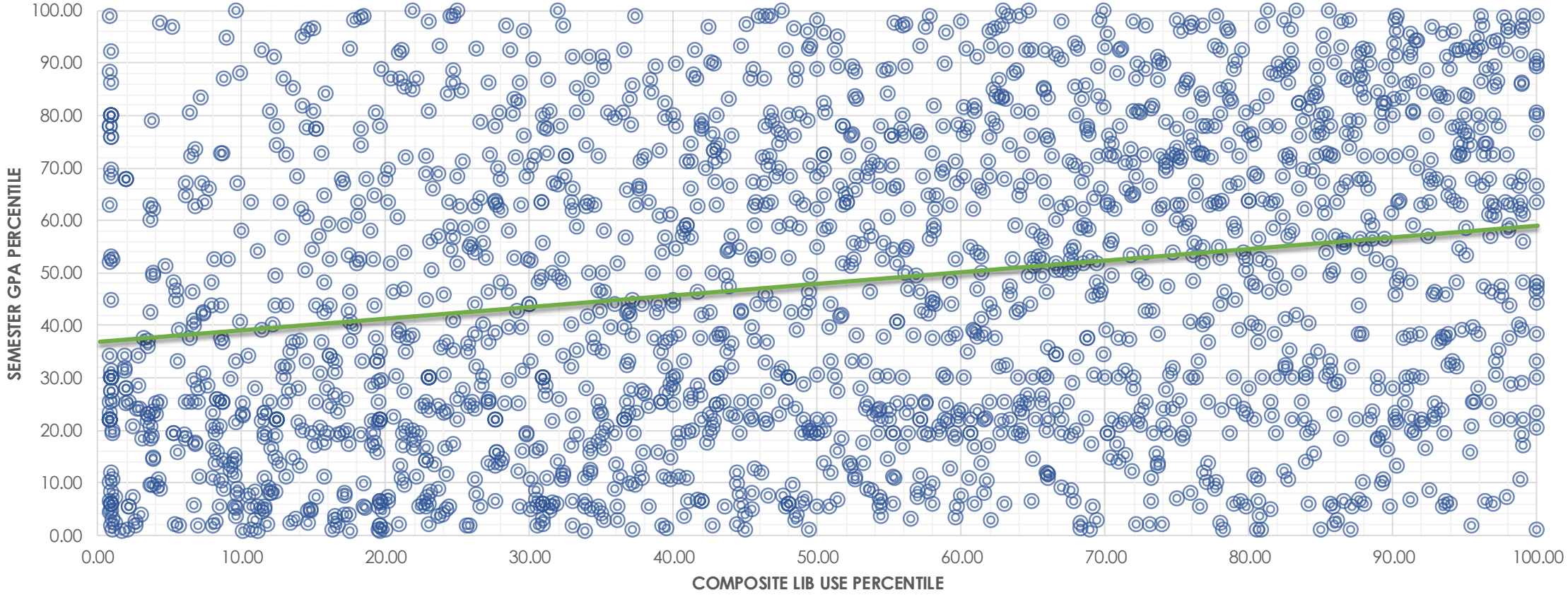
Q&A

GLM: Our Statistical Model



Q&A

A Quick Glance: Scatter Plot of Semester GPA Percentile x Composite Lib Use Percentile



Y = 36.3 + 0.22X

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.235[a]	.055	.055	27.80966	.055	1006.118	1	17182	<.001

a. Predictors: (Constant), LIB Percentile Rank for 5 Types (Loan/Entrance/E-Sessions/Booking/Workshops)

Q&A

Generalized Linear Mixed Model

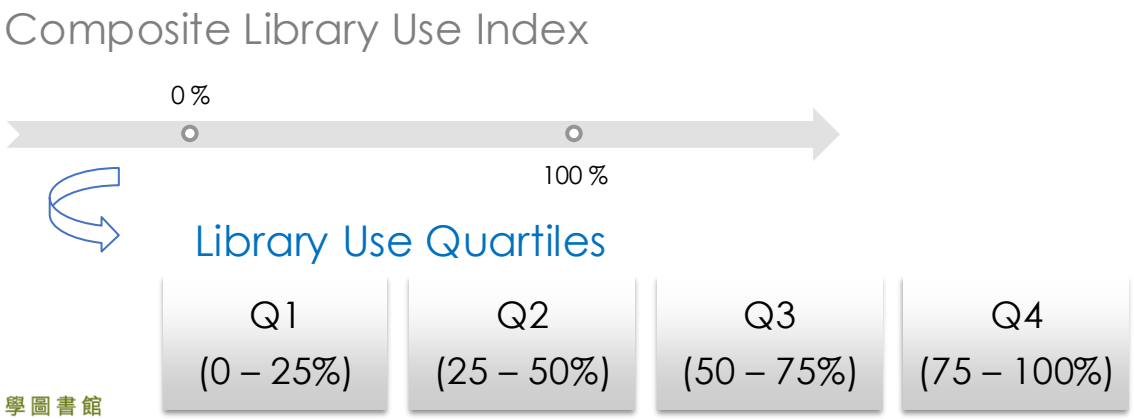
We chose Generalized Linear Mixed Model (GLM) as the model of analysis because:

- Multiple Independent Variables
  - We are trying to find out if we take other factors such as programme of study, gender, residency and hostel living status into account, will library use exhibit a positive relationship with student success?
- Repeated Measurements
  - GLM provides more flexibility over other models such as multiple linear regression model because it allows for repeated measurements. (we have GPA from the same student from 2 school terms)
  - GLM provides additional flexibility allowing us to treat different independent variables as fixed factors or random factors based on our model design.

Q&A

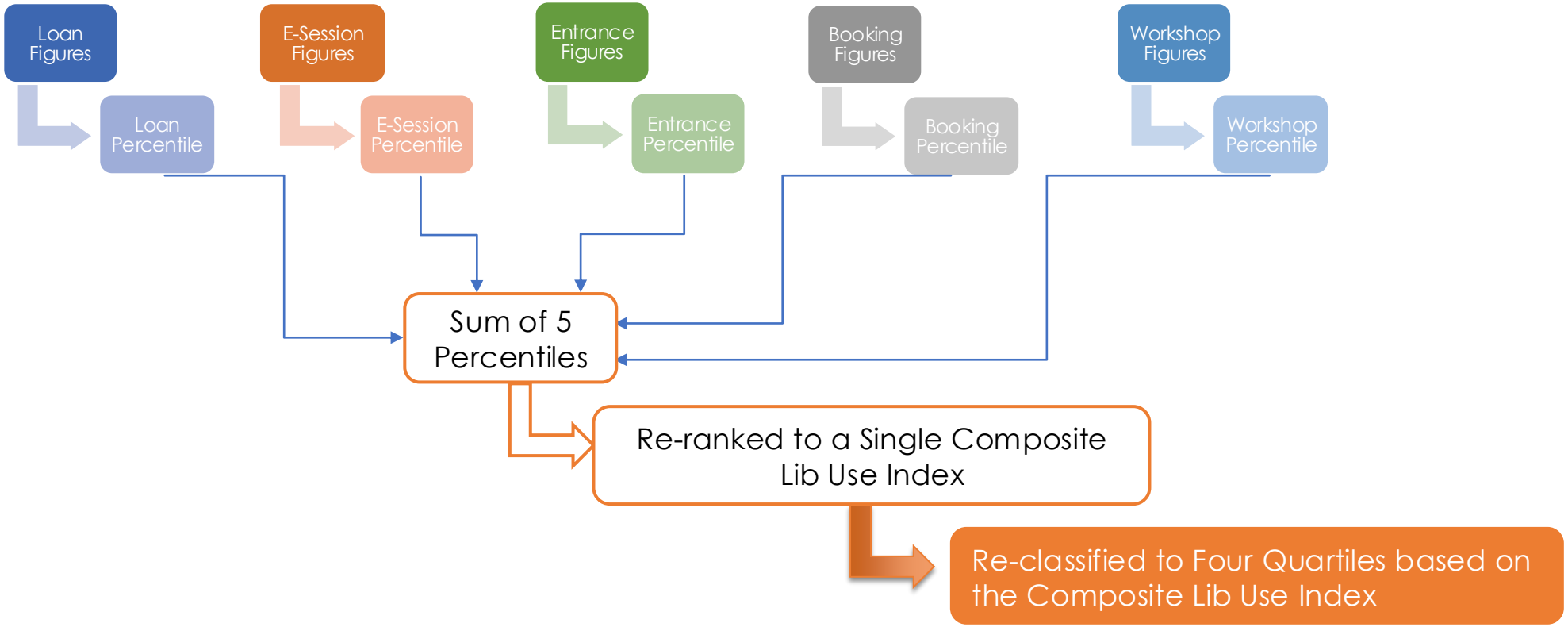
Final Data Transformation: Users Further Grouped into 4 Quartiles of Library Use Percentile Groups

- We grouped the students into 4 quartiles of library use so we can directly compare the strength of library use as a predictor of academic success against other fixed factors in our model, namely gender, hostel and residency status, which are also categorical variables.



Q&A

Further Library Use Data Transformation



Q&A

Our Statistical Model:

Statistical Module	Generalized Linear Mixed Model Module (SPSS)
Target Variable	Semester GPA Percentile
Fixed Factors	○ Lib Use (4 Quartiles) ○ Hostel Living Status (Yes / No) ○ Gender (Male / Female) ○ Residency (Local / Non-Local)
Random Factors	○ Programme of Study ○ Year of Study

Q&A

Results

Q&A

Results: Fixed Effects

Fixed Effects				
Source	F	df1	df2	Sig.
Corrected Model	176.256	6	17143	0.000
LIBUSE_4Q	296.979	3	17143	0.000
Hostel	0.828	1	17143	0.363
Gender	118.582	1	17143	0.000
Residency	12.560	1	17143	0.000

Note: these are results controlling for Programme of Study and Year of Study as random factors.

Significant at .05:

- The Statistical Model
- Library Use (4 Quartiles)
- Gender
- Residency

Not Significant:

- Living in Hostel or Not

Q&A

Results: Coefficients for Fixed Effects

Fixed Coefficients [a]

Model Term	Coefficient	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	57.257	1.3267	43.157	0.000	54.656	59.857
LibUse Q1	-17.137	0.5970	-28.704	0.000	-18.307	-15.966
LibUse Q2	-11.162	0.6138	-18.185	0.000	-12.365	-9.959
LibUse Q3	-6.175	0.6008	-10.278	0.000	-7.352	-4.997
LibUse Q4	0[b]					
Hostel: No	0.737	0.8107	0.910	0.363	-0.851	2.326
Hostel: Yes	0[b]					
Male	-5.593	0.5136	-10.890	0.000	-6.599	-4.586
Female	0[b]					
Non-Local	-2.243	0.6328	-3.544	0.000	-3.483	-1.002
Local	0[b]					

a. Target: Semester GPA in Percentile

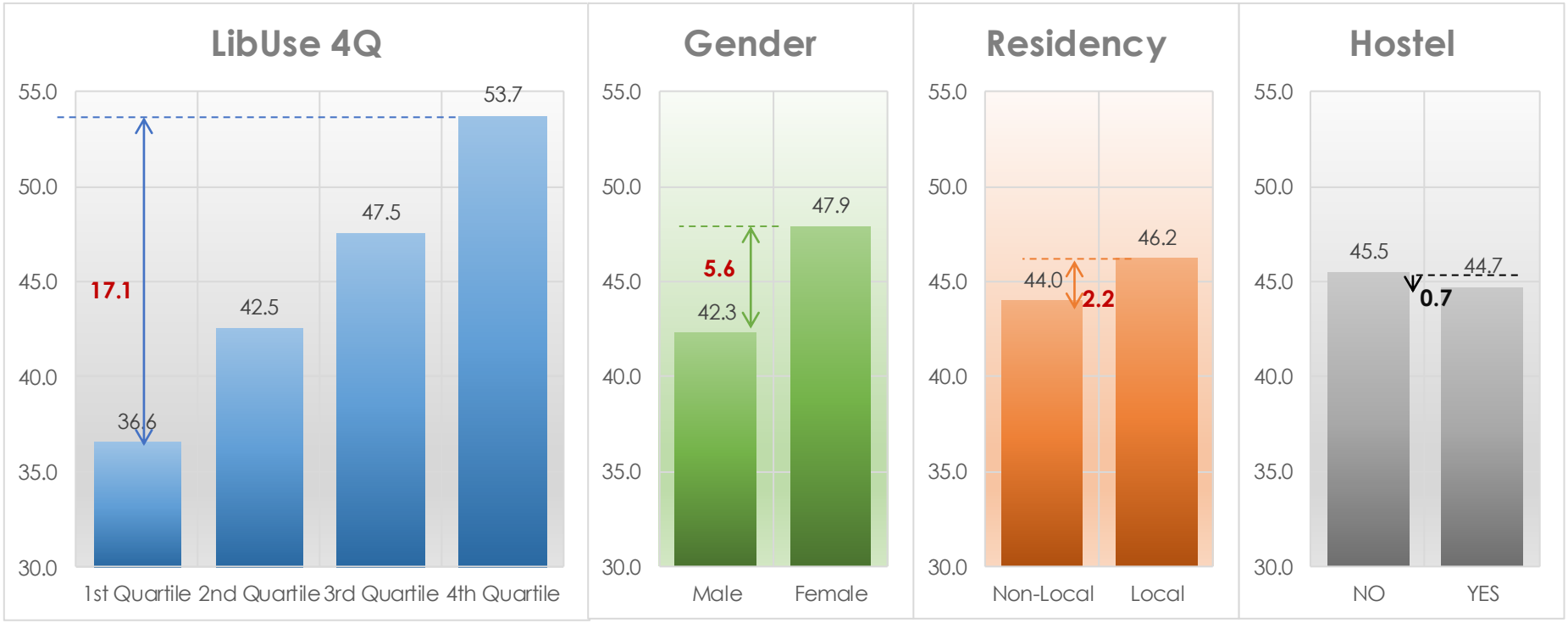
b. This coefficient is set to zero because it is redundant.

Interpretation:

- If we have to predict the average academic standing of a female local student who lives in the hostel and uses the library a lot (LibUse = top quartile), our best guess is 57.3 percentile.
- Everything else being equal:
  - LibUse Q4 → Q1, our best guess will be 40.2 percentile. (↓17.1%)
  - Female → Male, our best guess will be 51.7 percentile. (↓ 5.6%)
  - Local → Non-Local, our best guess will be 55.1 percentile. (↓ 2.2%)

Q&A

Estimated Means for Fixed Factors
Lib Use vs Gender vs Residency vs Hostel



Significant at .05 level? ✓ ✓ ✓ ✗

香港教育大學圖書館
EdUHK Library

Q&A

Detailed Tables
Lib Use vs Hostel vs Gender vs Residency

Estimates				
Lib Use 4 Quartiles	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
Q1	36.571	1.138	34.340	38.802
Q2	42.545	1.147	40.296	44.794
Q3	47.533	1.137	45.305	49.761
Q4	53.708	1.133	51.486	55.929

Q4 – Q1 = **17.1 Percentile** (**SIGNIFICANT**)

Estimates				
Gender	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
Male	42.293	1.131	40.076	44.509
Female	47.886	1.085	45.760	50.012

Female – Male = **5.6 Percentile** (**SIGNIFICANT**)

Estimates				
Living in Hostel	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
NO	45.458	1.002	43.493	47.423
YES	44.720	1.283	42.205	47.236

Hostel NO – YES = 0.7 Percentile (**NOT SIGNIFICANT**)

Estimates				
Residency	Mean	Std. Error	95% Confidence Interval	
			Lower	Upper
Non-Local	43.968	1.177	41.661	46.275
Local	46.211	1.067	44.119	48.302

Local – Non-Local = **2.2 Percentile** (**SIGNIFICANT**)

Q&A

Summing Up …

While we cannot infer causal relationship between library usage and student success, the statistical results have provided clear evidence that high library usage is indeed positively associated with high student GPA.

We also found evidence that when compared to gender, residency and hostel living status, library usage level serves as a better predictor of student GPA regardless of academic discipline and year of study.

Q&A

Thank you