

Chemistry Education Research and Practice

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Formative assessment by first-year Chemistry students as predictor of success in summative assessment at a South African university

Hlengani J Siweya ^{*a}, Peter Letsoalo ^b

Abstract

This study investigated whether formative assessment is a predictor of summative assessment in a university first-year Chemistry class. The sample comprised a total of 1 687 first-year Chemistry students chosen from the 2011 and 2012 cohorts. Both simple and multiple linear regression (SLR and MLR) techniques were applied to perform the primary aim of the research. In order to apply the above mentioned techniques, a selection criterion was executed on the 1 687 sample, after which 1 519 cases remained for the analysis. The study revealed a statistically significant SLR model, suggesting that formative assessment (FA) can, at 45.4% (that is $R^2 = 0.454$) level of accuracy, predict the summative assessment (SA) of students in a university first-year Chemistry class. Furthermore, the results of two MLR models discovered that SA can be predicted by using *theory marks* at 57.1% (that is $R^2 = 0.571$) level of accuracy, and *average semester test marks* at 59.4% (that is $R^2 = 0.594$) level of accuracy. The aforementioned domino effects suggest that the semester tests marks are more efficient, among other marks, in predicting the SA marks of students in a Chemistry department at a South African traditional university; accordingly recommends that more effort be made in preparing students for their semester marks.

In addition the study found that of the 1 519 students who had the 40% subminimum entry requirement for summative assessment, 765 (51.4%) passed the summative assessment, that of the 277 whose formative assessment mark was between 40% and 49% (both inclusive), 6 (2.2%) passed the summative assessment, that of the 1 208 whose formative assessment mark was at least 50%, 725 (60%) passed the summative assessment and, finally, that of the 34 students whose formative assessment mark was at least 75% only 8 were able to retain their grades.

Key Phrases: formative assessment and summative assessment

1. Introduction

It is often difficult to explain the discrepancy between the Formative and Summative Marks of many students at higher education institutions (HEIs). Research has shown that inconsistent standards of assessment in both formative and summative settings do lead to inconsistent performance by students at university; and that, on the relationship between formative and summative assessments, it has been shown that many students' performance in formative assessment (FA) far exceeds their performance in summative assessment (SA). There is also speculation that these discrepancies are attributable to a plethora of issues, chief amongst which could be: cheating by students in one form or other, repetition by lecturers of previous years' test questions or assignment tasks or tutorial test questions – from whence, without understanding of or insight into solutions, the students memorise marking schemes and reproduce them in the tests or examinations.

For these and other reasons, we compared formative and summative marks of first-year Chemistry students over a period of two years: from a sample of 1 519 first-year Chemistry students of the 2011 and 2012 cohorts at the university where the study was conducted, 37% and 32% of these respective cohorts had formative assessment marks of at least 50% but failed to acquire *the* subminimum¹ of 40% in the summative assessments (see Figure 1).

[Insert Figure 1 Here]

First-year Chemistry at *the* university consists of two modules which are basic requirements for all students who are registered for a first science degree in the Physical Sciences and Life Sciences streams, they are also prerequisites for most of the modules that are compulsory in this degree. In this connection, it would be interesting from an academic point of view on the one hand, and faculty management point of view on the other, to know whether or not the performance of first-year Chemistry students in formative assessment can be used to predict their performance in the summative assessment, as well as to know the correlation strength between students' formative

¹ A *subminimum* is the least mark a student needs to obtain in the summative assessment, in addition to the minimum 50% required as a final examination mark, to be considered to have passed the module.

1
2
3 assessments and their summative assessments in a Chemistry department at a
4 traditional² South African university.
5
6
7

8 9 **2. Background and Theoretical Framework**

10
11 Over the past few years, assessment has become a very important topic in both
12 secondary and tertiary education in South Africa to the extent that the performance of
13 students does not only become a record to be content with but forms an integral part of
14 the learning and teaching processes in schools and HEIs. In fact, there are HEIs (and few
15 secondary schools) where Teaching Excellence awards are dependent on teaching
16 portfolios and the pass rates of students in the subjects taught by the recipients.
17
18
19
20
21
22

23 **2.1 Formative and Summative Assessments**

24
25
26 There are many views, not necessarily convergent, on what *formative assessment* is.
27 Our understanding of formative assessment is as follows: a *weighted mark* obtained
28 from a collection of predetermined assessment tasks. These tasks may include
29 assignments, quizzes, projects and tests (Theory, Tutorials and or Practicals). An
30 assessment task is also considered to advance learning in so far as it provides
31 information that a lecturer can use in assessing his / her teaching effectiveness. This
32 form of assessment has been referred to as '*assessment for learning*' by Sorenson
33 (2000). Formative assessment should also assist students learn better upon receipt of
34 assessment feedback, which helps the lecturer assess himself or herself and helps
35 students assess themselves or one another. In agreement, Black and William (2004)
36 argue that "... assessment becomes *formative assessment* when the evidence is actually
37 used to adapt the teaching work to meet learning needs" and by improving performance
38 in formative assessments academic standards are raised as well. In addition, Crumrine
39 and Demers (2007) noted that "A useful arsenal of formative assessment tools becomes
40 nothing more than a cluster of gimmicks if not used in a way to inform instruction". In
41 effect, this implies that not only should formative assessment help the student in
42 exerting himself or herself in a module but "for assessment to be formative, it requires
43 feedback which indicates the existence of a 'gap' between the actual level of the work
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

² A traditional university (in South Africa) is one which offers theoretically-oriented university degrees.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

being assessed and the required standard. It also requires an indication of how the work can be improved to reach the required standard” (Taras, 2005). A similar view is echoed by Osado, Sánchez-Barba Merlo and Honrado Díaz del Campo (2013) in the sense that “formative assessment provides a collaborative framework that generates useful feedback information that can be used by both learners and teachers”.

In this study, we understand *summative assessment* to be taken at the end of the module, so it is expected to be based on and to cover all the topics outlined in a module: it is an *end-of-the* module examination. Unlike formative assessment, summative assessment is not used to inform learning by students but rather to inform the lecturer and the department whether or not the student has acquired enough knowledge to have passed the module and thereby declared *competent*. According to Taber (2003), summative assessments “are intended to test how well candidates know (can recall, apply and explain) their science”. In this study, we follow Ussher and Earl (2010) and Taras (2005) in that every “... process of assessment leads to summative assessment, that is, a judgement which encapsulates all the evidence up to a given point. This point is seen as a finality at the point of judgement”. For that reason, “summative assessments happen too far down the learning path to provide information” to the department for it to provide remedial interventions during the course of the teaching and learning of the module as is the case with formative assessment (see Garrison and Ehringhaus, 2007).

2.2 Topics covered in the modules

First-year Chemistry at this institution comprises two modules (or courses) taken over two semesters: a semester is a six month-long period starting in January or July and ending with semester examinations in May / June or November / December, respectively. In the first semester, students take General Chemistry 1A in which a student is considered competent if they are conversant with the terminology used in Chemistry, the periodic table, quantitative chemical relationships, thermo-chemistry, gases, colligative properties of solutions, acids and bases, chemical equilibria in acid or base systems, electrochemistry, the rates of reactions and thermodynamics

1
2
3 The second semester first-year Chemistry module students take General Chemistry
4 1B and their competence is judged on them proving competence in the Atomic and
5 Electronic Structure, chemical bonding, intermolecular attractions and liquids and
6 solids, alkanes, alkenes and alkynes, and alkyl Halides, alcohols, aldehydes and Ketones,
7 acids and derivatives, aromatics, amines and conformational analysis.
8
9
10
11
12

13 14 **3. Scope of the Study** 15 16

17 Teaching and Learning is a wide-ranging field, with a robust hypothetical and
18 experiential base (Biggs & Tang, 2011), so this study influences primarily the
19 assessment of teaching and learning in chemistry education. It also adds value to
20 reported studies that explain the relationship between FA and SA (see for instance,
21 Black *et al*, 2004; Sorenson, 2000; and Taras, 2005). The study aims to determine
22 whether or not the performance of first-year Chemistry students in FA can be used to
23 predict their performance in the SA in a Chemistry department at a South African
24 traditional university. (SA (Summative Assessment) is the final assessment of a module
25 that takes place at the end of a semester.) The study also investigated what the overall
26 pass rate was of students whose FA was at least 40%; determined whether or not the
27 students whose FA fell between 40% and 49%, both inclusive, passed the module after
28 their first ordinary examination; determined whether or not the students whose FA
29 ranged from 50% to 74% passed the module after writing their first ordinary
30 examination; as well as investigated whether or not those students whose FA were at
31 least 75% (considered to the least final examination mark required for a distinction)
32 retained their status during the first ordinary examination.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 **4. Methodology** 49 50

51 **4.1 Research settings and student population** 52

53 Marks for first-year Chemistry major modules for the 2011 and 2012 academic years,
54 obtained from the Department of Chemistry at the university where this study was
55 conducted, were used as secondary data for the analysis. These records were sourced
56 from the Chemistry lecturers offering the two modules through the Head of Department
57 and there was agreement that neither the names of the lecturers nor those of the
58
59
60

students enrolled for the two modules would be divulged. All variables included in the data are: FA marks with a weight of 100%; SA marks with a weight of 100%; average practical tests (AVRPT) marks with a weight of 50%; theory marks weighted at 50%; average semester tests (AVRSEMT) marks with a weight of 40%; average assignments and/or tutorial tests (AVRATT) marks with a weight of 10%. The following is the equation presenting how FA of each student was calculated:

$$FA = \text{Theory mark} + \text{AVRPT mark}$$

where $\text{Theory mark} = \text{AVRSEMT mark} + \text{AVRATT mark}$.

Although the data indicates how many semester tests were written it does not specify how many practical tests, assignments and tutorial tests were written by students (see **Table 1**). The data show only the AVRPT, average assignments and/or tutorial tests (AVRATT) for every student.

Table 1: Number of Assessments by year, module and semester

| Year | Semester | Number of Assessments | | |
|------|-----------------|-----------------------|-------|--------|
| | | Semester tests | AVRPT | AVRATT |
| 2011 | 1 st | 4 | 1 | 1 |
| | 2 nd | 4 | 1 | 1 |
| 2012 | 1 st | 3 | 1 | 1 |
| | 2 nd | 4 | 1 | 1 |

4.2 Data collection and selection criteria

In order to obtain the final data to perform the analysis, the selection criteria (SC) require the student to have

- a) written all semester tests;
- b) obtained at least a mark for AVRPT, obtained at least a mark for AVRATT and obtained a subminimum of 40%.

After implementing the selection criteria on the original secondary data with 1 687 registrations, the data remained with 1 519 registrations (see **Table 2**).

Table 2: The number of registrations before and after implementing SC

| Year | Semester | Number of Registrations (before SC) | Number of Registrations (after SC) |
|--------------|-----------------|-------------------------------------|------------------------------------|
| 2011 | 1 st | 430 | 379 |
| | 2 nd | 416 | 375 |
| 2012 | 1 st | 417 | 396 |
| | 2 nd | 424 | 369 |
| Total | | 1 687 | 1 519 |

5. Data Analysis and Results

5.1 Data examining

Descriptive Statistics was used on the data for the analysis (see **Table 3**). As can be seen from table, SA indicated a very high standard deviation (13.153) with the mean (42.78) that is below 50%; when compared with FA marks, SA shows a higher variation with a lower mean value. Consequently, this seems to suggest inconsistency between FA and SA. A 0* minimum mark of AVRATT does not necessarily mean the value zero but a mark between 0 and 0.5, both exclusive. Based on the mean of the Theory and AVRPT marks, seemingly these cohorts were performing better in the Practicals than in the Theory tests (that is, the semester tests and AVRATT).

Table 3: Descriptive Statistics

| Marks | N | Minimum | Maximum | Mean | Std. Deviation |
|---------------|-------|---------|---------|-------|----------------|
| AVRSEMT (40%) | 1 519 | 8 | 38 | 22.59 | 5.010 |
| AVRATT (10%) | 1 519 | 0* | 5 | 2.71 | 0.755 |
| Theory (50%) | 1 519 | 11 | 43 | 25.31 | 5.198 |
| AVRPT (50%) | 1 519 | 15 | 47 | 32.20 | 4.528 |
| FA (100%) | 1 519 | 40 | 85 | 57.25 | 8.157 |
| SA (100%) | 1 519 | 8 | 80 | 42.78 | 13.153 |

5.2 Regression and correlation analysis

Although the results of students of the same institution, class or course are positively correlated because they share several unobserved factors, the main statistical technique for making proper adjustments while measuring uncertainty is Regression Analysis. The data was taken from first-year Chemistry modules of 2011 and 2012 for 1st and 2nd semesters. These modules were taught by different lecturers, so the assumptions on variance-covariance structure that assume independence of observations will not be

violated (Matilde, *et al*, 2000). In order to investigate whether or not it is appropriate to use the first-year Chemistry students' FA to predict their SA in a Chemistry department at a South African traditional university, we employed both Simple and Multiple Linear Regression (SLR and MLR) analyses. These techniques are expected to produce a *prediction function* for forecasting the value of a response variable using predictor variable(s) (Graybill and Iyer, 1994). Another statistical method that was used is Correlation Analysis (CA); this will describe the relationship between the response and predictor variables in terms of both the strength of the relationship and the direction (see also Schwab, 2007 and Ho, 2006). From the dataset of 1 519, three groups were formed based on the predictor variable(s) as reflected in Table 4 below:

Table 4: Variables by Groups

| Groups | Response Variable | Predictor Variable/s |
|--------------|-------------------|---------------------------|
| Group 1 (G1) | SA | FA |
| Group 2 (G2) | SA | Theory and AVRPT |
| Group 3 (G3) | SA | AVRPT, AVRATT and AVRSEMT |

The datasets were then used to design 3 models all of which were performed at 95% confidence level. We designated them Model1 of G1, Model2 of G2 and Model3 of G3.

When performing this regression analysis technique, we performed regression diagnostics; where the relationships between the predictor variables and the response variable was found to be linear; the errors were also normally distributed; the error variance was found to be constant; the errors associated with one observation were not correlated with the errors of any other observation. Observations with large residuals, observation with an extreme value on a predictor variable, and influential³ observation were removed when performing regression diagnostics. The general rules of thumb that were used to investigate and omit the aforementioned observations are when the observation has:

$$\text{leverage} > \frac{2k+2}{n}; \text{abs}(rstu) > 2; \text{Cook's } D > \frac{4}{n}; \text{or } \text{abs}(DFBETA) > \frac{2}{\text{sqrt}(n)},$$

where k is number of predictors and n is the number of observations (IDRE, 2005).

³ Observation is said to be influential if removing the observation substantially changes the estimate of coefficients

5.2.1 Results and interpretation for Model1 of G1

After performing the regression diagnostics on the dataset of G1 with 1 519 registration, 141 observations were found to be outliers or influential and were then removed from the dataset of G1. All of the 141 observations were substantially different from all other observations; therefore there was a large improvement in the results of regression analysis after their removal. **Figure 2** shows the examined data before and after regression diagnostics in the form of a scatterplot, clearly indicating an improvement of the linear relationship.

[Insert Figure 2 Here]

Regression analysis for Model1 of G1 was performed with 1 378 registrations. The model contains the variable FA mark with a multiple R of 0.674, revealing a strong positive relationship between FA and SA. The R^2 of 0.454 ($R^2 = 0.674^2$) which means that the accuracy of predicting scores for the students' SA marks will be at approximately 45.4% if the prediction is based on scores for FA mark. The FA mark was found to be statistically significant with *p-value* less than 0.05 (see Sig. in Table 6); hence, for every additional *one* mark on the FA mark, the SA mark will increase by 0.992. The model that should be used to predict the SA at approximately 45.4% accuracy is Equation (1), where the FA mark of a student should be between 40% and 100%, both inclusive.

$$\widehat{SA} = -13.969 + 0.992(FA) \quad \dots (1)$$

Table 5: Model Summary for Model1 of G1

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|--------------|-------|----------|-------------------|----------------------------|
| Model1 of G1 | 0.674 | 0.454 | 0.454 | 8.922 |

Table 6: Coefficients for Model1 of G1

| Model1 of G1 | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|----------------|-----------------------------|------------|---------------------------|--------|------|
| | B | Std. Error | Beta | | |
| Constant | -13.969 | 1.696 | | -8.238 | .000 |
| Formative Mark | 0.992 | 0.029 | 0.674 | 33.869 | .000 |

5.2.2 Results and interpretation for Model2 of G2

In order to build a regression model that includes more than one predictor variable the stepwise procedure was followed, where the model includes only variables that are significant at 95% confidence level. The regression diagnostics test was implemented on 1 519 observations; which proposed that 186 observations should be removed in order to improve the linear relationship of the variables and also to increase the accuracy of the model, and indeed we then removed the observations. The dataset of G2 was then left with 1 333 on which Theory and AVRPT marks (as constituent parts of FA) were used to investigate whether or not they can predict SA. In this model, the partial-regression plots were very useful in identifying influential points. The relationship between SA and Theory seem to have improved after regression diagnostics test whereas the relationship between SA and AVRPT still remained somewhat the same.

[Insert Figure 3 Here]

From the test results we deduced that only the Theory mark is significant at 0.05 significant level with a multiple R of 0.756; thereby revealing a strong positive relationship between Theory and SA marks. The R^2 of 0.571 which primarily means that the accuracy of predicting scores for the students' SA marks will improve by approximately 57.1% if the prediction is based on scores for Theory marks. The Theory marks were found to be statistically significant with a coefficient value of 1.7; thus inferring that for every *one* added mark on the Theory mark, the SA mark will increase by 0.995. The test revealed that the coefficient of the average Practicals mark is not significantly different from zero; so, it is not necessary to include this mark in the model. The model (*prediction function*) that should be used to predict the SA at approximately 57.1% accuracy is Equation (2), where Theory mark of a student cannot exceed 50%.

$$\widehat{SA} = -1.175 + 1.7(\text{Theory}) \quad \dots (2)$$

Table 7: Model Summary for Model2 of G2

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|--------------|-------|----------|-------------------|----------------------------|
| Model2 of G2 | 0.756 | 0.571 | 0.571 | 7.781 |

Table 8: Coefficients for Model2 of G2

| Model2 of G2 | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|--------------|-----------------------------|------------|---------------------------|--------|-------|
| | B | Std. Error | Beta | | |
| Constant | -1.175 | 1.047 | | -1.123 | 0.262 |
| Theory | 1.700 | 0.040 | 0.756 | 42.105 | 0.000 |

5.2.3 Results and interpretation for Model3 of G3

The partial-regression plots were also very convenient in identifying influential points in Model3. In **Figure 4** the linear relationship appears to be positive between SA and all predictor variables. Although before the regression diagnostic was performed the scatterplots had shown to have some outliers and influential observations, after removing them the linear relationship between SA and AVRSEMT improved considerably as compared to others. Ostensibly, the two predictor variables AVRPT and AVRATT might not be significant in predicting the SA marks of the students.

[Insert Figure 4 Here]

Remember that Theory is the summation of AVRSEMT and AVRATT. After the use of three predictor variables in the dataset of G3 and performing the regression diagnostics (which left the dataset of G3 with 1 316 observations) the model contained only one significant variables (AVRSEMT) with a multiple R of 0.771. This part of the analysis also used the stepwise procedure in the MLR analysis technique; hence the R^2 of 0.594, suggesting that the accuracy of predicting scores for the students' SA marks will be approximately 59.4% if the prediction is based on scores for AVRSEMT only. The coefficient of 1.797 for AVRSEMT mark means that for every additional *one* mark on the AVRSEMT mark, the SA mark will increase by 1.797 when the AVRATT mark is held constant. It was also observed that the coefficient of the AVRATT and AVRPT are not significantly different from zero at 95% confidence level, implying that the two variables should be removed from the model. The final model (*prediction function*) that can be used to predict the SA at approximately 59.4% accuracy is Equation (3), where AVRSEMT cannot exceed 45%:

$$\widehat{SA} = 1.306 + 1.797 (\text{AVRSEMT}) \quad \dots (3)$$

Table 9: Model Summary for Model3 of G3

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|--------------|-------|----------|-------------------|----------------------------|
| Model3 of G3 | 0.771 | 0.594 | 0.594 | 7.554 |

Table 10: Coefficients for Model3 of G3

| Model3 of G3 | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|--------------|-----------------------------|------------|---------------------------|--------|-------|
| | B | Std. Error | Beta | | |
| Constant | 1.306 | 0.952 | | 1.372 | 0.170 |
| AVRSEMT | 1.797 | 0.041 | 0.771 | 43.876 | 0.000 |

5.2.4 Models Comparison

R-squared only measures how close the data are to the fitted regression line; therefore it cannot determine whether or not the coefficient estimates and predictions are biased, it does not indicate whether or not a regression model is adequate. It is possible to have a low R-squared value for a good model, or a high R-squared value for a model that does not fit the data. This is why residual plots are used to assess the biasness of coefficient estimates and predictions; and the adequacy of a model.

In Figures 5, 6 and 7, the scatterplot of regression residuals and regression predicted value, show to have a rectangular pattern that deduce a symmetrical pattern and that have a constant spread throughout the range. It is also observed that there is not much difference between the points on the scatterplot; however this might mean that the best model in this case is Model 3 of G3 since it has the highest $R^2 = 0.594$.

[Insert Figure 5, 6, and 7]

5.3 Probability analysis

Probability is one other interesting analysis that has been a subject of discourse amongst philosophers, logicians, mathematicians, statisticians, physicists, and psychologists for many years (Bartoszynsk & Niewiadomska-Buga, 2007). Some scholars (see for instance, Keeney, 2009 and Ghahramani, 2005) insinuated that probability is a number between 0 and 1 that is associated with an event that is intended to represent its chance of occurring. *Conditional probability* is referred to as

1
2
3 the probability of one event if another event occurred (Carmen, 2007). In this study it is
4 denoted by $P(\text{Results status}/\text{population of interest})$, where results status can be
5 either *fail* (F), *supplementary* (S), *pass* (P) or *pass with distinction* (PD). The population
6 of interest was obtained by categorising the students into Category 1 (C1) of 1 519
7 students who have obtained $FA \geq 40\%$; Category 2 (C2) of 277 students who have
8 obtained $40\% \leq FA \leq 49\%$; Category 3 (C3) of 1 208 students who have obtained $50\% \leq$
9 $FA \leq 74\%$; and Category 4 (C4) with 34 students who have obtained $FA \geq 75\%$.

10
11
12
13
14
15
16
17 The results reveal that the chances of a student to failing; supplementing; passing
18 and passing with distinction given that the student has qualified to sit for SA, is 0.24;
19 0.25; 0.5 and 0.01, respectively out of 1. It has also exposed that approximately half of
20 the students who have obtained $FA \geq 40\%$ have a chance of passing the module.
21
22
23
24

25 The results indicated that, when given that a student falls in C2 the probability of the
26 student passing are 0.22 out of 1; proposing high chances (0.82 out of 1) of failing the
27 module; $P(S/C2) = 0.16$; and surprisingly $P(PD/C2)$ is zero out of 1. The students in C2
28 obtained the same or even lower SA marks when compared with the FA marks. Given
29 that a student has obtained formative mark between 50% and 74% the probability of
30 the student passing is 0.6, and this leaves more than quarter chances (0.4) of the
31 remaining students from C3 to sit for supplementary or fail. The students falling in C4
32 have approximately 0.24 chances out of 1 to maintain a distinction final mark; of course,
33 this also shows that the student from C4 has no chance of failing the module.
34
35
36
37
38
39
40
41
42

43 6. Discussion (and Implications)

44
45
46 This study lends support to Harlen's findings (2005), namely that there is need for
47 "synergy between formative and summative assessment requires that systems should
48 be designed with these two purposes in mind and should include arrangements for
49 using evidence for both purposes". The study has confirmed that, to a certain extent,
50 formative assessment does predict summative assessment. The fact that the majority of
51 those who entered summative assessment with formative assessment marks of at least
52 50% passed the summative assessment, this department of Chemistry has every reason
53 to invest resources in their first-year chemistry students to assist them get at least 50%
54 as a formative mark because then they know such students will pass the module.
55
56
57
58
59
60

6.1 Discussion and implications based on the designed models

Model diagnostics methods were performed with some outlying observations detected, which were then removed from the data and the models were then fitted. All three models were found to be adequate and the best model, that is, Model3 of G3, was chosen based on the prediction accuracy.

- When using only FA marks as a predictor variable for SA marks, the FA marks was found to be significant at 0.05 confidence level with an R^2 of 0.454, and a positive coefficient; hence Model2 of G2 has prediction accuracy of 57.1%. *Model1 of G1 suggests that formative marks of students can indeed be used to predict the summative marks of students, taking in to account that the prediction accuracy is 45.6%. However, a prediction accuracy of 45.6% is not satisfactory enough for the department to rely on.*
- When using both Theory marks and Practicals mark as predictor variables for SA marks, only Theory marks were found to be significant at 0.05 confidence level with an R^2 of 0.571, and a positive coefficient; hence Model2 of G2 has prediction accuracy of 57.1%. *In this particular model Theory marks contribute 50% of the FA mark; a greater concern is on the rejected variable, average practicals marks which contribute similar weight as the Theory mark. A weak relationship explored between the average Practical marks and SA marks points to the fact that summative assessments do not include module material that is done during practicals classes – and accordingly average Practical marks should not be used to predict SA marks of students in a 1st year Chemistry class.*
- When using both average semester test marks, average assignments and/or tutorial tests and Practicals mark as predictor variables for SA marks, Only average semester test marks was found to be significant at 0.05 confidence level with an R^2 of 0.594 and a positive coefficient; hence Model3 of G3 has prediction accuracy of 59.4%. *Based on the same reason why average Practical marks should not be used to predict the in Model 2 of G2, Model 3 of G3 indicates that only average semester tests marks can be used to predict SA marks of students in a 1st year Chemistry class with nearly 60% level of accuracy.*

1
2
3 It is certainly clear that Model3 of G3 is the best model than the other two to use in
4 predicting the SA marks of students in a 1st year Chemistry class.
5
6
7
8
9

10 **6.2 Discussion and implications based on the probability analysis**

11
12 This analysis was performed in order to know students' chances of obtaining
13 different results status given the category (based on FA mark) where the student falls.
14
15
16

17
18 Students with a formative mark between 40% and 49% (inclusive) are highly likely,
19 $P(F/C2) = 0.82$, to fail the module; hence the department of chemistry should consider
20 arranging extra tutorials or providing assistance to students at risk or even other
21 student interested during the study week. This might also reduce the chances of
22 supplementary in C2 and C3. A continuation of a 0.22:1 ratio for a student in C2 to pass
23 the module, might lead to a determination of setting a minimum mark out of 100% for
24 student to qualify to write summative assessment to 50%.
25
26
27
28
29
30
31
32

33 **6.3 Conclusion**

34
35 The findings have important implications for heads of department, and directors of
36 school; even moreso, it can be useful to individual course instructors and students.
37 However, the study implicitly questions the standard that the lecturers apply during FA
38 against that used during SA. In view of the theoretical framework followed in the study,
39 the authors are concerned that there did not seem to have any assessment feedback in
40 any of the tests (theory and practicals) administered – which begs the question: what is
41 the lecturers' understanding of formative assessment? Inevitably these questions need
42 further investigation if the faculty in which these modules are administered wishes to
43 intervene and improve the students' performance. The department has a responsibility
44 to advocate for a balance between *assessment of learning* and *assessment for learning* –
45 both of which require assessment feedback strategy if students are to benefit from any
46 form of assessment and for the department to be confident that its academic offerings
47 make the desired impact.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Notes
4
5

6 * Corresponding author
7

8
9 ^aExecutive Dean, Faculty of Science and Agriculture, University of Limpopo, Private Bag
10 X1106, Sovenga, 0727, South Africa. Email: Hlengani.Siweya@ul.ac.za
11

12
13 ²Faculty Researcher, Faculty of Science and Agriculture, University of Limpopo, Private
14 Bag X1106, Sovenga, 0727, South Africa. Email: Peter.Letsoalo@ul.ac.za
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

7. References

- Bartoszynsk, R., & Niewiadomska-Buga, M. (2007). *Probability and Statistical Inference* (2nd ed.). New Jersey: John Wiley & Sons.
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University* (4th ed.). New York: Two Penn Plaza.
- Black P. and William D., (1998), "Inside the Black Box: Raising Standards Through Classroom assessment, *Phi Delta Kappan*, pp. 139-48.
- Black P., Harrison C., Lee C., Marshall B. and William D., (2004), Assessment for Learning in the Classroom, *Phi Delta Kappan*, **86** (1), pp. 8-21.
- Crumrine T., & Demers C. (2007) Formative assessment: Redirecting the plan. *The Science Teacher*, 74(6), 64-68.
- Garrison C. and Ehringhaus M., (2007), *Formative and Summative Assessments in the Classroom*, Association for Middle Level Education, August 2013.
- Ghahramani, S. (2005). *Fundamentals of Probability with Stochastic Processes* (3rd ed.), London: Pearson Education.
- Graybill, A. F., & Iyer, K. H. (1994). *Regression Analysis: Concepts and Applications* (1st ed.). Duxbury Pr.
- Harlen W., (2005), Teachers' summative practices and assessment for learning – tensions and synergies, *The Curriculum Journal*, **16** (2), pp. 207 – 223.
- Ho, R. (2006). *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Boca Raton: Taylor & Francis Group.
- IDRE, (2005). *Institute for Digital Research and Education, Statistics, University of California*, <http://www.ats.ucla.edu/stat/spss/webbooks/reg/chapter2/spssreg2.htm>, Accessed December 11, 2013.
- Matilde, B., Paola, M., Domenico, P., & Luigi, S. (2009). *Statistical Methods for the Evaluation of Educational Services and Quality of Products*. London: Springer.

- 1
2
3
4 Osado A. G, Sánchez-Barba Merlo L.F., Honrado Díaz del Campo M., (2013), Formative
5 Assessment in Chemistry Course through feedback and feed-forward strategies,
6 Proceedings of the 7th International Technology, Education and Development
7 Conference, 3852-3860.
8
9
10
11 Schwab, A. James. Course Materials. 2007. <http://www.utexas.edu/courses/schwab/>
12 (accessed June 01, 2013).
13
14
15
16 Sorenson, P. (2000) Drowning in numbers? The need for formative assessment. In *Issues*
17 *in science teaching*, ed. Sears, J. and Sorenson, P. pp. 123–132. London: Routledge
18 Falmer.
19
20
21
22 Taber K. S., (2003), Examining structure and context – questioning the nature and
23 purpose of summative assessment, *School Science Review*, **85**(311), 35-41
24
25
26
27 Taras, M., (2005), *Assessment – summative and formative – some theoretical reflections*,
28 *British Journal of Educational Studies*, ISSN 0007-1005, Vol. 55 (4), 466 – 478.
29
30
31
32 Ussher, B. & Earl, K., (2010), ‘Summative’ and ‘Formative’: Confused by the Assessment
33 Terms?, *New Zealand Journal of Teachers’ Work*, Volume 7, Issue 1, 53-63.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

8. Figures

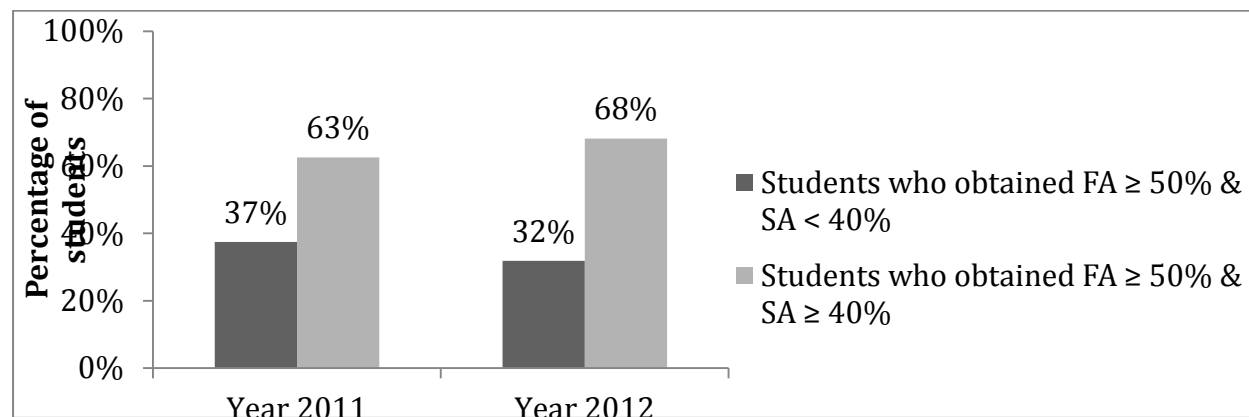


Figure 1: Students who obtained FA ≥ 50% & SA < 40% and FA ≥ 50% & SA ≥ 40%

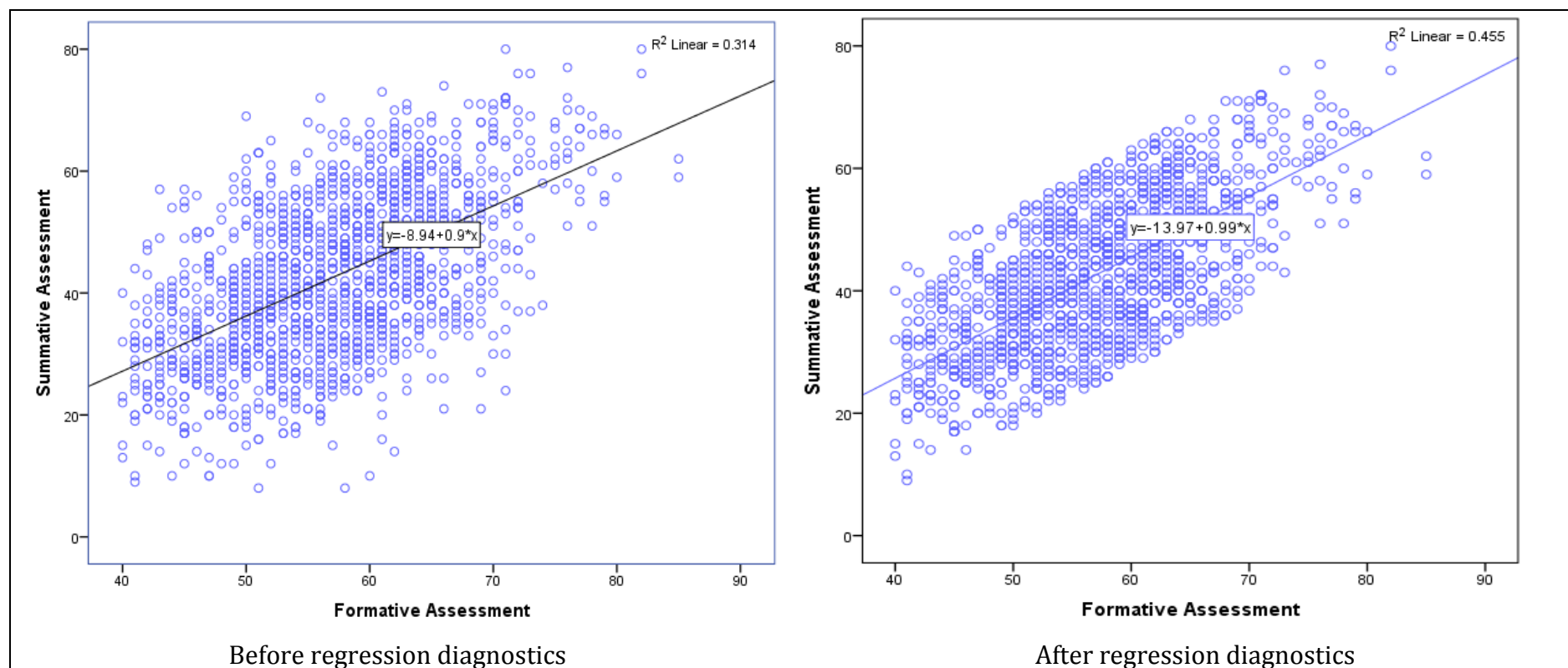


Figure 2: Scatterplot of SA by FA

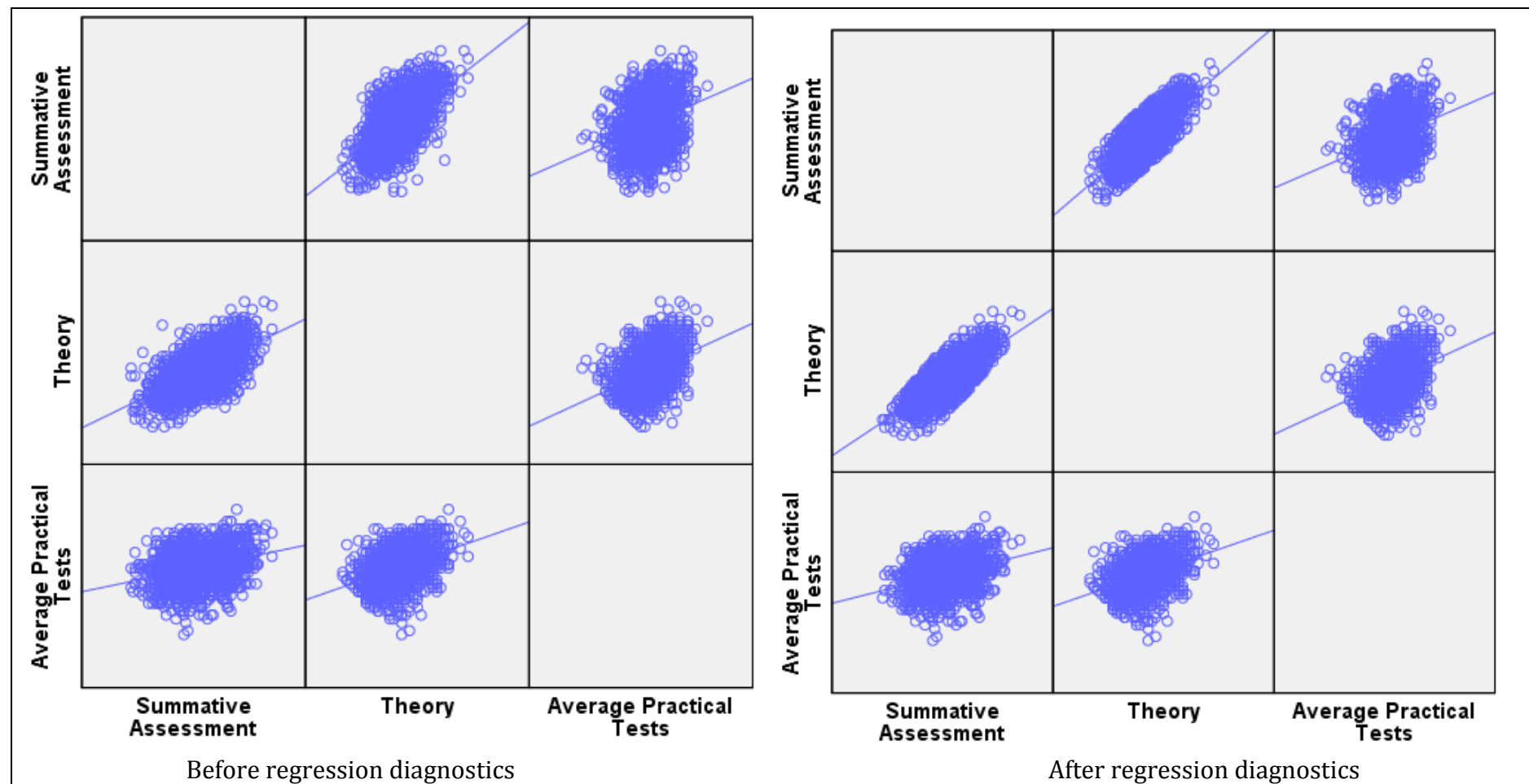
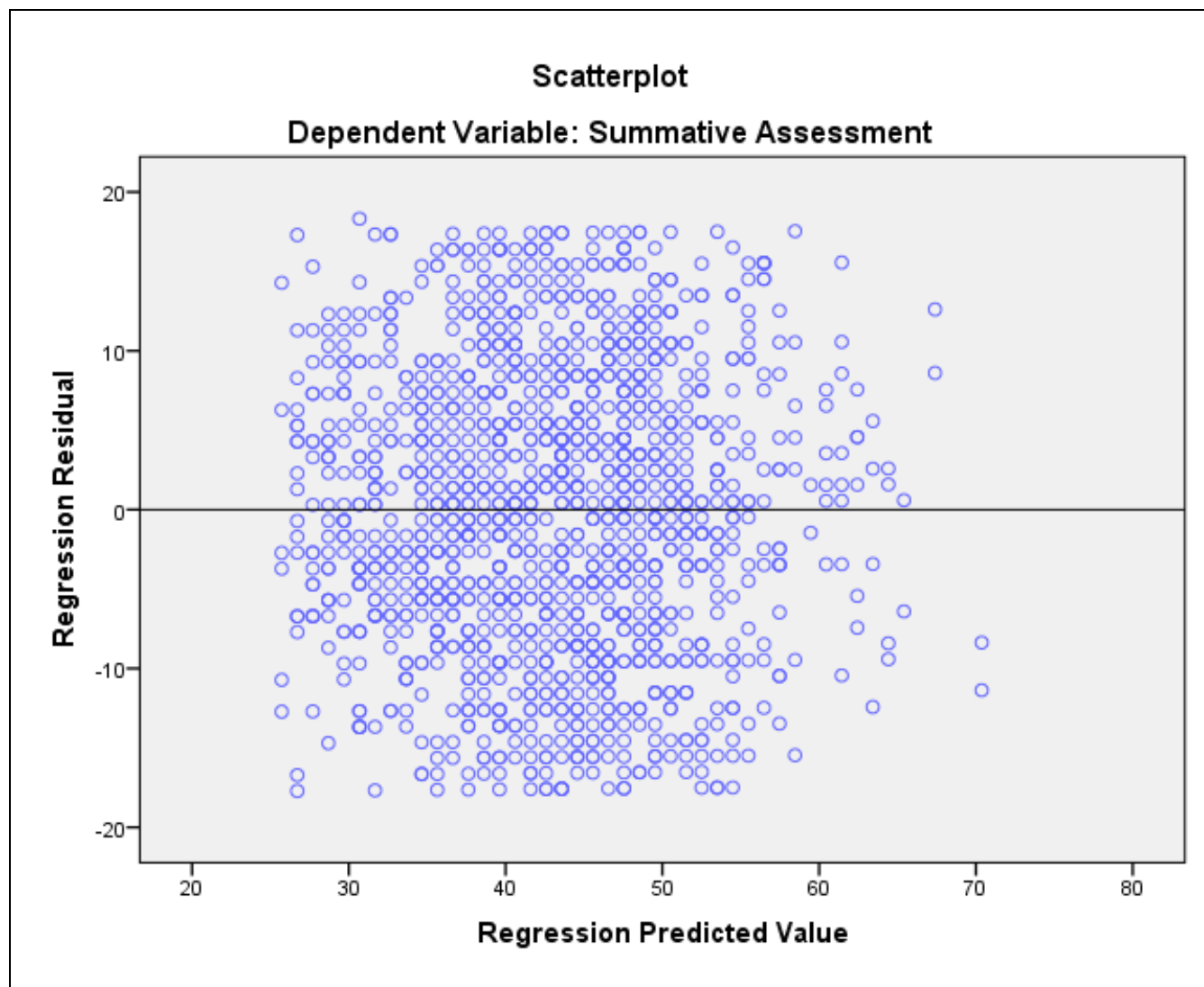


Figure 3: Matrix Scatterplot of SA by Theory, AVRPT



39
40
41
42
43
44
45
46
47

Figure 5: Residual plot for Model 1 of G1

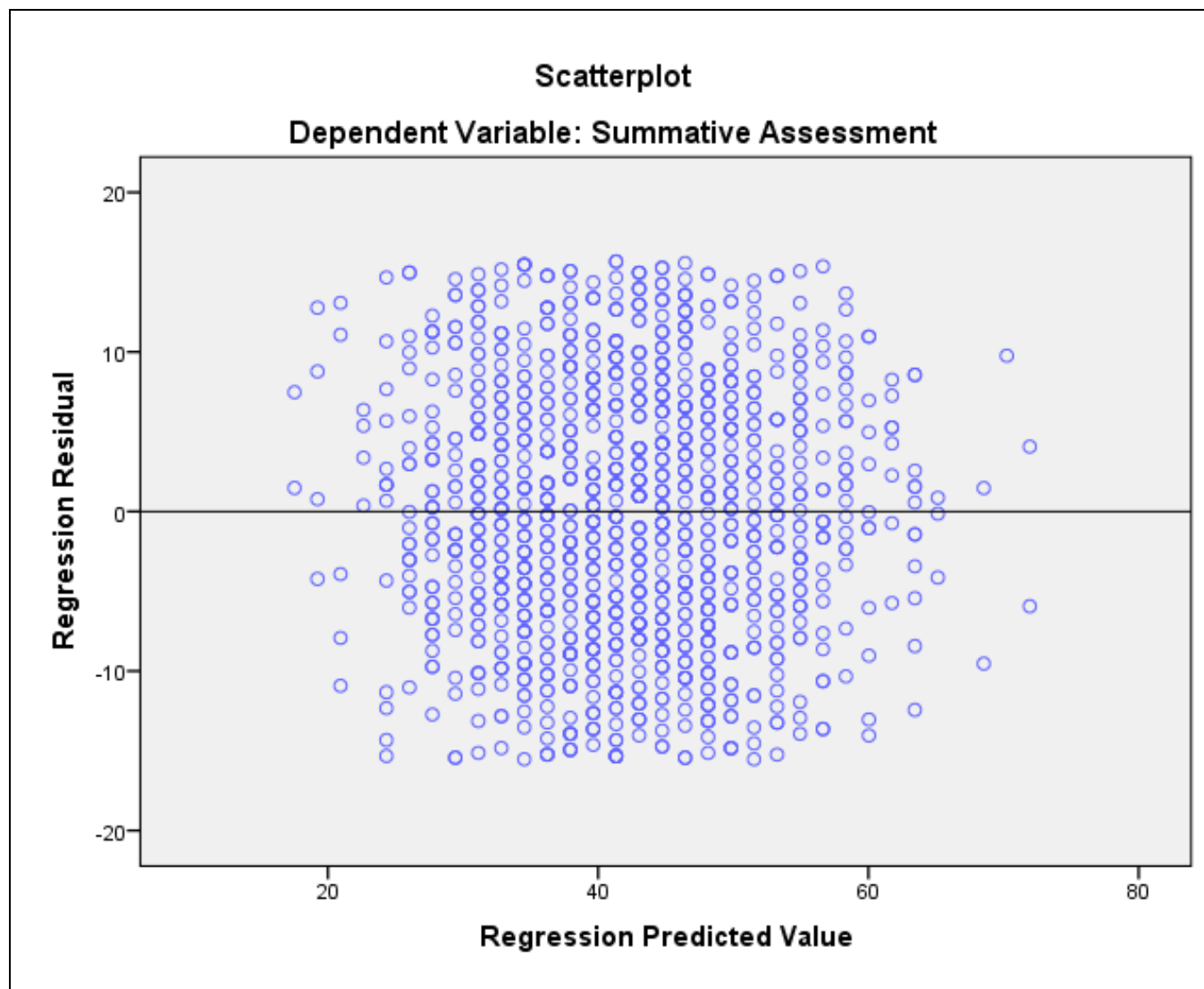
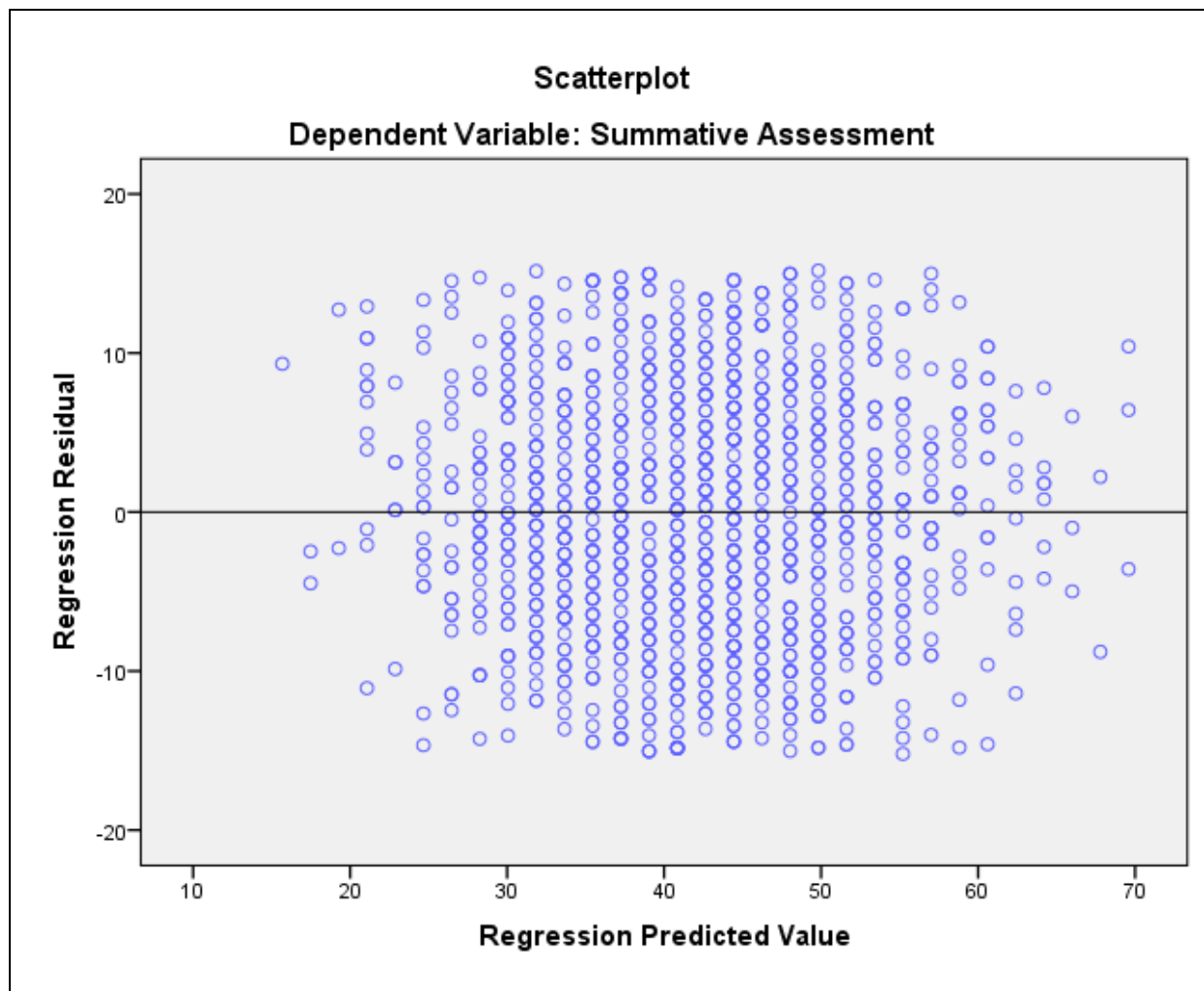


Figure 6: Residual plot for Model 2 of G2



39
40
41
42
43
44
45
46
47

Figure 7: Residual plot for Model 3 of G3

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47