

CHAPTER THREE

PILOT STUDY

3.1 Introduction

A Pilot study was undertaken to determine the reliability of the psychological scales used and the validity of each of the scale items. Altogether, the questionnaire used in the pilot had seven scales. The seven scales were the identity development, self-esteem, social responsibility, attitude towards authority, social skills effectiveness, goal-setting and locus of control scales. The identity development scale was developed by the researcher, while the other scales were modifications of existing scales.

In the first pilot undertaken, all the six established scales attained acceptable levels of reliability. The reliability of the identity development scale (developed by the researcher), however, was too low to be acceptable. The identity development scale was thus redesigned and tested by means of a second pilot survey, where an acceptable reliability level was achieved.

This chapter describes firstly how the data was collected for the first pilot, the design of the identity scale by the researcher and the reliability and validity results of the identity scale from the first pilot study. This is followed by an explanation of how the identity development scale was redesigned for the second pilot. Next, the

second pilot is described. This is followed by the reliability and validity results of the identity scale in the second pilot test.

Thereafter, the preparation and the translation of the other 6 scales, which are the self-esteem, social responsibility, attitude towards authority, social skills effectiveness, goal-setting and locus of control scales are explained. Additionally, the reliability of the scales as well as the validity of each of the scale items are described.

3.2 Subjects

For the pilot test a convenience sample was used. The students were from two schools, that is, La Salle Brickfields and Methodist Boys School, Sentul. The students were all from Form 5 Science Two. The second best class, was chosen to have a sample as close as possible towards the average cognitive ability of students. Further, they were all males, aged 17 years old. In each of the classes, all the students who attended class on the day of the survey participated in the survey. Unfortunately however, on the day of the survey, there were seven absentees from Methodist Boys School and one absentee from La Salle Brickfields. They were excluded from the survey. In total, 32 students from La Salle Brickfields and 28 students from Methodist Boys School participated in the survey.

The two schools may be considered comparable, because they were both urban schools with average students, that is to say, they were not elite schools. Thus,

students from both schools were possibly similar to each other in cognitive ability and also came from similar socio-economic backgrounds.

3.3 Procedure

The school headmaster had allowed the researcher the use of a double-period, that is 80 minutes, to undertake the survey. The students were briefed on how to complete the questionnaire. The students were told that this was not a test and that there were no right or wrong answers. The contents of the questionnaire were explained. The researcher explained section by section, giving details of the possible choices of answers and the necessity to choose only one answer that reflected, or that most nearly reflected their views or opinions. Further in each section, one question was discussed as an example to give a clearer picture of how to answer the questions.

The survey questionnaire was self-administered by the students. The researcher was present throughout the survey. Any inquiries were immediately answered. The students had no difficulty in answering the questions. They took 30 - 45 minutes to complete the questionnaire. The questionnaires were then collected and analysed.

3.4 Questionnaire Development

The questionnaire consists of 4 four sections. The first section which was made up of 74 items measured identity status, self-esteem, social responsibility and attitude

towards authority. This section was scored on a 5 point Likert scale with answers ranging from “strongly disagree” to “strongly agree”. The second section consisted of 12 items and measured social skills. This section was also scored on a 5-point Likert scale. However, the answers ranged from “always admire” to “always dislike”. The third section contained 12 items and measured goal-setting. It was scored on a 7-point semantic differential scale. The final section which measured locus of control had 30 items. It was scored on a forced-choice format.

3.4.1 Identity Development Scale

3.4.1.1 Scale Development

Firstly, the identity status level of the respondents needed to be ascertained. According to Marcia’s model there were four status levels, that is, achievement, moratorium, foreclosure and diffusion.

To determine the identity status level, Marcia (1966) undertook 15 - 30 minutes semi-structured interviews. All interviews followed the same outline, although deviations from the standard form were permitted in order to explore areas more thoroughly. The interview was terminated when all the questions in the outline were completed as well as the interviewer felt that he had enough information to categorise the individual. Interviews were tape-recorded and replayed for judging. Often each interview was heard twice, sometimes three or four times.

Marcia had prepared a scoring manual to be used by each of the experimenters. Each subject was evaluated in terms of the presence or absence of crises as well as the degree of commitment. Three domains were investigated, that is, occupation, religion and politics. The interview judge familiarised himself with the descriptions of the statuses provided in the Manual and then sorted each individual according to the categories in the manual based on the pattern he most closely resembled.

A sample question in the occupational area was:

How willing do you think you'd be to give up going into _____ if something better came along? Examples of typical answers for the four statuses were:

(Identity Achievement) Well, I might, but I doubt it. I can't see what "something better" would be for me.

(Moratorium) I guess if I knew for sure I could answer that better. It would have to be something in the general area ----- something related.

(Foreclosure) Not very willing. It's what I have always wanted to do. The folks are happy with it and so am I.

(Identity Diffusion) Oh sure. If something better came along, I'd change just like that.

A sample question in the religious area was:

Have you ever had any doubts about your religious beliefs?

(Identity Achievement) Yeah, I even started wondering whether or not there was a God. I've pretty much resolved that by now, though. The way it seems to me is

(Moratorium) Yes, I guess I'm going through that now. I just don't see how there can be God and yet so much evil in the world or

(Foreclosure) No, not really, our family is pretty much in agreement on these things.

(Identity Diffusion) Oh, I don't know, I guess so. But it doesn't really bother me much. I figure one's about as good as the other.

It was felt that Marcia's instrument would be difficult to use as the respondent's answers have to be fairly specific and narrow to be able to be classified in the proper status level. Too wide an answer would make categorisations difficult. Further, the answers, even the above answers in the samples given by Marcia, can be interpreted differently by different people. The interpretation was too subjective. Another problem was that it would be difficult to use this technique on a large sample.

It was thus felt that an objective scale would be more effective if the respondent chose answers that clearly and more specifically indicated his own status level. Further, the scale should be a multiple choice format to ensure that the respondent chose only one specific answer, from the options given. His choice would indicate his

status level. Additionally, it would use the paper and pencil method, which could easily be employed on a large sample.

Dellas and Jernigan (1987) used the Dellas Identity Status Inventory-Occupation (DISI-O) to assess occupational identity. It was a forced-choice, objectively scored, paper-and-pencil instrument. Based on statements measuring the presence or absence of crises and commitment, the DISI-O instrument produced separate meaningful factors representing 5 sub-scales, that is, achievement, moratorium, foreclosure, diffused-diffused and diffused luck. According to Dellas and Jernigan diffused-diffused was when there was no commitment and only superficial search while diffused-luck meant no commitment and dependence on luck or fate.

The DISI-O instrument consists of seven sets of statements. Each set contains five statements and each statement represents an identity status. Subjects select one statement that is MOST LIKE ME from each of the seven sets. The selection of four or more statements representing a particular identity status classifies an individual as a member of that status. An individual who does not select at least four statements on a particular status was considered Unclassified. Dellas and Jernigan only researched occupational identity.

Streitmatter (1988) in his study on ethnicity and identity development used the Extended Objective Measure of Identity Status (EOM-EIS) instrument to assess the

identity status of his early adolescent samples. The instrument was an objective self-reporting measure adapted from the interview format constructed by Marcia (1966). It again examined 4 specific stages of psychosocial maturity, that is, identity achievement, moratorium, foreclosure and diffusion. It is a 124 item instrument which measured the domains of occupation, politics, religion, interpersonal relationships, dating, sex roles and recreation.

While the instrument used by Streitmatter may have been useful for the research being conducted, the researcher was unsuccessful in obtaining a copy of the instrument.

Thus, based on the principles of the objective scales used above a new instrument was developed to determine Identity Status level. The scale needed to be objective and a paper-and-pencil measure to identify the identity status level of respondents. Marcia had used the domains of career, politics and religion in his study of identity status. Streitmatter apart from those three domains, included the domains of interpersonal relationship, dating, sex roles and recreation.

The domains selected to be used to determine the identity status in the current study were career, religion, interpersonal relationships, politics and sports and recreation. These domains were selected because it was felt that they would be the most relevant for Malaysian adolescents. Adolescence is a stage where there is

increasing concern for one's future and one's career path. There is also increasing religious and political awareness and questioning to better understand one's religion and also political processes such as issues of freedom and democracy. Further, adolescence is a stage where one begins to interact with a wider array of persons in school, community, work-place or youth clubs and thus interpersonal skills are important for one's personal and social development. Further, the use of leisure time for sports and recreation also continues to be of interest at this stage. The domains of dating and sex roles were dropped as it was felt that Malaysian adolescents are still not actively concerned of this component of life. The original design of the scale consisted of forty four items.

There were five sub-scales as follows:

Items 1-8	Career status	8 items
Items 9-16	Religious status	8 items
Items 17-32	Interpersonal status	16 items
Items 33-36	Political status	4 items
Items 37-44	Sports and Recreation status	8 items

For each item, the respondent had to choose from a five-point Likert scale,

that is:

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

The statements were arranged in sets of 4 items, that is, the first item reflected an achievement status, the second item a moratorium status, the third item a foreclosure status and the fourth item the diffusion status. The sequence of the four statuses was maintained throughout the questionnaire. For example, the first 4 items were related to career. The first item was "After exploring several possibilities, I am now clear as to what I want to do for my future career" (achievement). The second item was "Until now I am still unclear about my capabilities and what occupation that is most suited to me" (moratorium). The third item was "My parents have decided what is best for my career and I will abide by their decision" (foreclosure). The fourth item was "I have still not made any career plans. For now, anything will do as long as there is an opportunity" (diffusion).

As the differences in meanings between each of the four items were not very precise, it was hoped that after completing the first two or three sets of four items the respondent would have an understanding of the response choices for the rest of the instrument. This would possibly enable consistent scoring. If the items were not in any particular order, it was feared that the close meaning of the statements would confuse the respondents. While there was a danger of respondents selecting an automatic response, it was felt the risk of confusing the participants by changing the item sequence outweighed the risk of automatic response.

The respondents were to choose the position on the scale that represented or most nearly represented their position. The status of the respondents was to be determined by analysing the scores. For each sub-category, the status level selected most frequently would indicate the status of the individual. For example, if in the career sub-scale, the respondent chose the achievement items most, through selecting strongly agree or agree, he would be classified as identity achievement status. It was expected that there would be a high level of consistency in the answers chosen which would indicate the status level of the respondent.

3.4.1.2 Results

The Pilot test provided extremely disappointing results. Overall reliability was insignificant with an r-value of only 0.06. Further, out of the 44 items, only 9 were found to be valid. The summary of the results is as follows:

Table 3.1. Summary of the Test of Validity for the Identity Scale

Status	No. of Items	Valid	p-values	Not Valid
Career	8	1	1 item = 0.01	7
Religion	8	2	2 items= 0.01	6
Interpersonal	16	4	2 items= 0.01 2 items= 0.001	12
Political	4	0	-	4
Sports and Recreation	8	2	2 items=0.001	6
Summary	44	9		35

Several reasons for the extremely low reliability are suggested. Firstly, the differences between the items may not have been sufficiently clear, that is the meaning of items may have been too closely related. Respondents may not have been able to discern the differences between the different items.

Further, it had been assumed that respondents would score consistently based on their status levels. Thus, of the 4 possible statements, each indicating a particular status level, the respondent would consistently chose one, that is "agree" or "strongly agree" for the level he was at and "unsure" or "disagree" for the other levels. This, however, was not the case in the results. Respondents often chose similar answers, such as "agree" or "strongly agree", for different status level. Thus determining the status level of the respondent became extremely difficult. It was clear that the current format was not appropriate.

3.4.1.3 Scale Redesign

It was decided to redesign the Scale and carry out a second pilot, just for the identity status measure.

Firstly, the format of the answers was restructured. In the new format, each item now carried four choices, from which the respondents had to choose one. Each of the answers reflected a status level, that is, the first answer was that of achievement status, the second one of moratorium, the third one of foreclosure and the fourth one

of diffusion status. The respondent was expected to choose the answer that represented or most nearly represented his position or view. Thus for each item, the answer selected indicated his status level. For example, the first item stated, "With reference to my future career/occupation:"; the four choices given were, "After exploring various possibilities, I have made a decision" (achievement), "I am still in the process of exploring various possibilities, before making my final decision" (moratorium), "My parents/relatives know better the best decision, and I will abide by their choice" (foreclosure), "I have not made any decision and for me it is really irrelevant to make any choice" (diffusion). Thus for each item, the respondent had to choose one out of 4 choices. That choice was assumed to represent a status level.

The number of items were also expanded on. In the earlier format, 4 items had to be scored consistently for just one aspect of the category. For example, in the earlier format, career status had 8 items. The eight items dealt with career choice. A foreclosure person for example would have to choose "strongly agree" or "agree" for foreclosure only and not for any other items (in the 8 items on career) in the first 4 sets of items as well as the second 4 sets of items. For all the rest 6 items, he must choose the other choices or at least must not choose "strongly agree" or "agree" consistently for another status level over the 2 sets of 4 items. Thus out of 8 items on career, 2 items had to be scored specifically on that particular status, while the rest 6 items could be freely scored as long as no other statuses were also consistently scored "strongly agree" or "agree" over the two set. For example if a respondent also chose

“strongly agree” or “agree” for moratorium in the first set of 4 items as well as second set of 4 items, it would be impossible to categorise the respondent. Thus it can be said that although 8 items measured career status, only 2 items out of 8 were actually taken into account.

In the newer format, since each statement was represented by only one item, the number of statements could be expanded on to capture more effectively the status level. Thus in the previous questionnaire there were 2 statements in the career category (represented by 8 items), 2 statements in the religious category (represented by 8 items), 4 statements in the interpersonal category (represented by 16 items), 1 statement in the political category (represented by 4 items) and 2 statements in the sports and recreation category (represented by 8 items). These were changed in the new questionnaire to 5 items in the career category, 5 items in the politics category, 4 items in the religious category and 7 items in the interpersonal status category. In summary, the difference between the previous and new questionnaire was that in the previous questionnaire there were 36 items (not including sports and recreation), which measured the response of 9 statements where else in the new questionnaire, there were 21 items that measured the response of 21 statements.

The new format with each item representing one statement allowed more items for each category. Further, the format of the answers where the respondent chooses one answer from 4 answers, where each answer indicated a particular status level,

allowed unambiguous scoring. Thus a clearer answering format enabled respondents to score more directly while more items for each category enabled a more accurate discrimination of the status level.

In the previous format, the respondent had to consistently score over sets of 8 items (in career), before his status level was determined. Thus he had to select his level and reject the other levels. This approach while more difficult for the participants, nevertheless meant a generally stronger scale. In the new format, the respondent needs to select one out of 4 response choices, thereby making the task easier as well as clearer. However, by not having to indicate their position on each status, it will be more difficult to assess the consistency of answers within an item. Consistency will now be assessed across items of the scale.

While it would probably be easier for the respondents to choose from the response choices, this format risks the possibility that respondents may select the response choice that gives the preferable account of themselves, which may or may not be a true account of themselves. The researcher will be unable to discern the consistency of scoring to determine the status levels. This time the respondent chooses his status level, but does not have to consistently reject the other status levels. This generally makes for a weaker scale compared to the previous design. However as an initial measure, it was felt that an unambiguous response format was necessary as the first step in developing an effective measure of identity status.

Although there were 2 valid items out of 8 items in the earlier questionnaire, the sports and recreation sub-scale was dropped as it was felt that adolescents do not really give much serious thought to sports and recreation activities. For example, the item "I follow the recreational activities of my parents" (foreclosure) or the item "I am not interested in sports, as it is very boring" (diffusion), are not very suitable as adolescents do not give serious consideration to matters relating to sports and recreation in the context of determining personal identity. As such it was felt, that this domain may be perceived as contributing very little to the identity development of adolescents. This sub-scale was thus dropped.

The new questionnaire contained one section with 21 items which measured identity status. Of the 21 items, 5 items measured career status, 5 items measured political status, 4 items measured religious status while 7 items measured interpersonal status. For each statement, the respondent had to choose one of 4 choices, each answer indicated a particular status level. The sequence of the 4 answers was maintained throughout the questionnaire because it was felt that once students answered the first two or three questions, they may not read the answers for the later statements, as the wording of the answers would be similar and they would probably assume that the format was similar for the rest of the questionnaire. This sequencing would face the risk of respondents providing desired instead of actual answers. However, as the distinction between choices was fairly fine the sequence was maintained to minimise confusion in selection of the choices.

3.4.1.4 Second Pilot

A second pilot test, just for the new identity status measure was undertaken. A convenience sample of 67 students participated in the survey. The sample was made up of 40 students from Form 5 Science 2 in St. Johns and 27 first year tutorial students in University Malaya.

All students in Form 5 Science Two class were surveyed. All respondents were males aged 16-17 years. In the University Malaya tutorial class, all the students who had attended that particular class in which the survey was being undertaken were asked to complete the questionnaire. The total attendance for that class was actually 35 students. However, on the day of the survey only 27 had attended that tutorial. The age range of the class was 20-23 years old.

In both cases, the researcher gave instructions on how to answer the questionnaire. He explained that there were no right answers, and that for each statement the respondent had to choose only one of the four answers given. He further illustrated by answering one statement that was given in the questionnaire. The researcher was present throughout the survey. There were no inquiries as the students did not appear to have any difficulties. The questionnaire was completed in less than 15 minutes.

The two samples may not be entirely comparable as one, that is from St. Johns, were younger than the first year University students. However it was felt that the difference would not severely affect the analysis.

3.4.1.5 Results of the Second Pilot

In the second Pilot test, the results were more encouraging. The overall reliability was 0.68 at a p-value of 0.001 indicating the scale was reliable.

Table 3.2 Validity of Identity Scale Items

Item No.	R-Value	Item No.	R-Value
<u>Career</u>		<u>Religion</u>	
1.	0.52**	11.	0.19
2.	0.45**	12.	0.27
3.	0.30	13.	0.39**
4.	0.33*	14.	0.19
5.	0.25		
<u>Politics</u>		<u>Interpersonal</u>	
6.	0.26	15.	0.45**
7.	0.33*	16.	0.41**
8.	0.21	17.	0.38*
9.	0.42**	18.	0.38*
10.	0.27	19.	0.52**
		20.	0.49**
		21.	0.50**

p value * significant at 0.01 ** significant at 0.001

Out of the 21 items, 13 were found to be valid, that is 9 items had p-value of 0.001 and 4 items at p-value of 0.01. Eight items were not valid. Of the 8 non-valid items the lowest r-value was 0.19. The non-valid items were however included in the final questionnaire so that each sub-scale would not have too few items. Adding non-

valid items in the scale may reduce the strength of the scale. However, this was preferred to having too few items which would clearly affect the validity of the scales being measured.

One of the items, that is item number 21 which read, "in choosing my clothes", was dropped, although it had a p-value of 0.001 because it was felt that it was not consistent with the other items in the interpersonal sub-scale. The other items measured self-awareness, self-confidence, personal goals and choosing friends.

The final instrument used in the study consisted of 20 items as follows:

Items 1 - 5	Career status	5 items
Items 6 -10	Political status	5 items
Items 11 - 14	Religious status	4 items
Items 15 -20	Interpersonal status	6 items

3.4.2 Translation of the Scales

The following scales, that is the Self-Esteem Scale (Rosenberg, 1965a), the Social Responsibility Scale (Berkowitz and Lutterman, 1968), Value Profile Scale (Bales and Couch, 1969), Personal Value Scales (Scott, 1965) and Purpose-in-Life (Crumbaugh, 1968) were taken and modified from Robinson and Shaver's (1973) Measures of Social Psychological Attitudes. In the text, the scales were in the English Language

and needed to be translated to Bahasa Malaysia before the actual research could be undertaken.

The procedure used for translation was through back-translation (Brislin, 1970). In this process, the scales in English were translated to Bahasa Malaysia and another, also bilingual colleague, translated the scales back to the English Language. This process was repeated a second time by two different bilinguals. The final English Language text was compared to the original English Language, and since in all cases the meaning was judged to be similar, the final translated Bahasa Malaysia scales were used in the research.

3.4.3 Self-Esteem

Self-esteem in the study was measured by the Self-Esteem Scale (Rosenberg, 1965a). The original scale consisted of a 10 items on a four point scale, from “strongly agree” to “strongly disagree”. The Scale was changed to become a 5-point scale, from “strongly agree” to “strongly disagree”. This was done as it was felt that a 5-point scale captures the response more effectively than a 4-point scale. Further, a 5-point scale is the scale commonly used by researchers. Further, it was decided to standardise all scales in the questionnaire to a 5-point scale.

Table 3.3 Summary of the Validity of the Self-esteem Scale

Scale	No. of Items	Valid	p-value	Not Valid
Self-esteem	10	9	9 items = 0.001	1

In the Pilot, of the ten items, 9 were found to be valid and one not valid. The item that was not valid was, "I certainly feel useless at times". As the item was ambiguous and difficult to interpret, the item was dropped. The 9 remaining items were used in the final study.

In the Pilot, the reliability test was undertaken using the split-half technique. Overall reliability was 0.59 at a p-value of 0.001. The scale was thus reliable.

3.4.4 Social Responsibility

Social Responsibility was measured by the Social Responsibility Scale (Berkowitz and Lutterman, 1968). The scale attempts to assess a person's social responsibility, that is an orientation towards helping others even when there is nothing to be gained from it. The original scale was on a 5 point interval scale, that is, "strongly agree" to "strongly disagree", and this was maintained.

Further, the original scale had 10 items. 2 items were dropped as it was felt that they were inappropriate in the Malaysian context. The two dropped items were, "It is no use worrying about current events or public affairs; I can't do anything about them anyway" and "Our country would be a lot better off if we didn't have so many

elections and people didn't have to vote so often". The first item assumes politically conscious youths while the second item refers to "too many elections", when in fact in Malaysia there is only one election every 5 years. The second item may be more suited in the West where there are far more elections at various levels, from city council to federal elections.

After removing two items, it was also decided to add four more items that reflected social responsibility in the local context. These new items were proposed as it was felt that local youths are more sensitive to concerns about underprivileged people as well as serving the community and nation. The four new items were, "We should be sensitive to the needs of the underprivileged", "Community work ought to be a part of school activities", "In this materialistic world, every one should take care of himself only", and "Each one of us, should whenever possible, serve the community and the nation".

Table 3.4 Summary of the Validity of the Social Responsibility Scale

Scale	No. of Items	Valid	p-value	Not Valid
Social Responsibility	10	6	6 items = 0.001	4

Of the 10 items, 4 items were not valid and they were dropped. The remaining 6 items were used in the final study.

In the Pilot test, the overall reliability of the scale was 0.72 at the significance level of 0.001. The scale was thus reliable.

3.4.5 Attitude Towards Authority

To measure Attitude towards Authority, the Value Profile (Bales and Couch, 1969) was used. The Scale measured 4 values, that is, acceptance of authority, needs determined expression versus value determined restraint, equalitarianism and individualism.

For the purpose of this study, only the Attitude Towards Authority measure was used. It consisted of 10 items to be scored in a 6 point scale. Consistent with the rest of the scales, a 5 point scale was used.

Table 3.5 Summary of the Validity of the Attitude Towards Authority Scale

Scale	No. of Items	Valid	p-value	Not Valid
Attitude Towards Authority	10	6	6 items = 0.001	4

Of the 10 items, 4 were found not to be valid and were dropped. The final scale used the 6 valid items

In the Pilot based, a reliability score of 0.69 was obtained. on the significance level of 0.001 The scale was thus reliable.

3.4.6 Social Skills

The Social Skills Variable was measured by the Personal Values Scale (Scott, 1965). The Scale was used to measure 12 personal values, that is, intellectualism, kindness, social skills, loyalty, academic achievement, physical development, status leadership, honesty, religiousness, self-control, creativity and independence.

In the social skills scale, there were 12 items. These 12 items were scored on a 3 point scale indicating: "Always admire," "Depends on Situation" and "Always Dislike". Consistent with the rest of the scales, again a 5-point scale was used.

Table 3.6 Summary of the Validity of the Social Skills Scale

Scale	No. of Items	Valid	p-value	Not Valid
Social Skills	12	9	7 items = 0.001 2 items = 0.01	3

Of the 12 items, three items were found to be not valid and were dropped. The 9 items remaining were used in the final scale.

The overall reliability score was 0.60 with a significance level of 0.01. The scale was thus accepted as reliable.

3.4.7 Goal Setting

Goal-setting was measured by the Purpose-in-life (PIL) Test (Crumbaugh, 1968). This scale was designed to measure the degree to which a person experiences a sense of meaning and purpose in life.

The PIL scale is made up of 20 items rated from 1 (low purpose) to 7 (high purpose). Total scores therefore ranged from 20 (low purpose) to 140 (high purpose).

Of the 20 items, 8 were dropped as it was felt that they were not appropriate in the local context or were ambiguous. The items were , "I am usually; completely bored (low purpose) to completely enthusiastic (high purpose)". This item was dropped as its meaning was similar to the next item when translated into Bahasa Malaysia. The item "If I could choose; I would prefer never to have been born (low purpose) to live nine more lives just like this one (high purpose)" was dropped because the concept of "nine more lives" is an unfamiliar expression in the local context. The item, "After retiring; I would do some of the exciting things I have always wanted to do (high purpose) to loaf completely the rest of my life (low purpose)", was dropped as it was felt that students have yet to consider issues of personal retirement. The item, "As I view the world in relation to my life; the world completely confuses me (low purpose) to fits meaningfully with my life (high purpose)" and the item "Concerning man's freedom to make his own choices I believe man is; absolutely free to make all choices (high purpose) to completely bound by limitations of heredity and

environment (low purpose)", was dropped because adolescents may have difficulty in understanding these concepts. The item, "I am a very irresponsible person (low purpose) to very responsible person (high purpose)" was dropped because it was felt that this concept was too direct and personal and may cause bias towards being a responsible person. The items, "With regard to death I am; prepared and unafraid (high purpose) to unprepared and frightened (low purpose)", and "With regard to suicide; I have thought of it seriously as a way out (low purpose) to never given it a second thought (high purpose)", was dropped as it was felt that student may not be concerned about issues of death and suicide. The remaining 12 items were used in the Pilot study.

Table 3.7 Summary of the Validity of the Goal-setting Scale

Scale	No. of Items	Valid	p-value	Not Valid
Goal Setting	12	11	8 items = 0.001 3 items = 0.01	1

Of the 12 items, one item was found to be not valid and dropped. The 11 valid items were used in the final scale.

The overall reliability score was 0.71 at a significant level of 0.001. This scale was thus accepted as reliable.

3.4.8 Locus of Control

Locus of Control was measured by the Rotter's Internal-External Control (IAR) Scale. The IAR Scale is a 34 item forced choice scale. Of the 34 items, 2 were dropped as they posed difficulties in translation. The dropped items were, "Suppose you weren't sure about the answer to a question your teacher asked you, but your answer turned out to be right. Is it likely to happen: ____ (a). because she wasn't particular as usual, or ____ (b). because you gave the best answer you could think of?" and "Suppose you weren't sure about the answer to a question your teacher asked you, but your answer turned out to be wrong. Is it likely to happen: ____ (a). because she was more particular than usual, or ____ (b). because you answered too quickly?".

Based on the point-biserial statistic, the r-value for $p=0.05$ was 0.25, on a two-tailed test. The results from the study showed that only 14 of the 32 items were valid. The remaining 18 items were not valid.

Additionally, the reliability score was also discouraging. A split-half analysis indicated that the scale was not reliable. The r-value was only 0.07.

It was thus decided that this scale was unreliable and thus cannot be used in the field study.

3.5 Conclusion

For the 6 scales selected, all the scales were reliable at p-value of 0.001. For the identity score, 13 items were valid and 7 were not valid. However, as dropping the non-valid items would seriously effect the number of items on each sub-scale, that is resulting in too few items, it was decided to include the non-valid items. This may probably slightly weaken the measurement of the scale, however, it was felt that this was preferable to having too few items for each sub-scale.

For the other scales however, all non-valid items from the pilot were dropped, and only the valid items were used.

Thus the final questionnaire had the following format:

Table 3.8 Contents of the Final Questionnaire

Scales	No. of Items	Reliability: r-value	p-value
Identity Development Score	20 items	0.68	0.001
Self- Esteem Score	9 items	0.59	0.001
Social Responsibility Score	6 items	0.72	0.001
Attitude towards Authority Score	6 items	0.69	0.001
Social Skills Effectiveness Score	9 items	0.60	0.001
Goal-setting Score	11 items	0.71	0.001