Bull. Nat'l Taiwan Nor. Univ.
Vol. 36, pp. 219-238 (1991)

師大學報
第36期 第219-238頁(民國80年)

CONSTRUCTION AND VALIDATION OF INSTRUMENT TO ASSESS CHEMISTRY MANIPULATIVE TECHNIQUE - SUBLIMATION

Cheng-Hsia Wang and Jhy-Ming Horng

Department of Chemistry

Abstract

This article describes a process to construct and validate criterion-referenced mastery measurement (CRMM) for sublimation technique. The results of this work are as follows: 1. Concept analysis was used in the analysis of sublimation to construct a better evaluation tool which gave accurate grades and searched out what mistakes the learner made so that the mistakes could be corrected by remedial instruction. 2. Content analysis and evaluation analysis were used to establish criteria which led to construction of an evaluation tool to assess the learning levels attained. 3. Reliability and validity of the tool are satisfactory for assessing whether the students can perform sublimation of ordinary cases in introductory organic chemistry laboratory course. 4. The designed process containing S-P chart analysis, a non-parametric method, is satisfactory in validating the assessment tool for classroom CRMM, as applied to sublimation.

INTRODUCTION

The current teachers' assessments of students' chemistry manipulative techniques are largely based on the impression grading during laboratory experiments and on degree of students' success at the end of a task, if available. Such assessments have thses weak points: (1) teachers can not give accurate grades; (2) the assessments can not search out what mistakes the learners made to allow remedial instructions; and (3) some students may perform as instructed, however, without assurance of learning transfer into a new situation. Since criterion-referenced mastery measurement (CRMM) is designed to find out which learning tasks a student can or cannot perform, rather than to discriminate among students, traditional index of item difficulty, index of item discriminating power, and correlation coefficient are of little value in determining reliability and validity of the instrument (CRMM). Furthermore, necessary data usually available are only from classes of about 20-50 students. A set of items in CRMM, for example, might be answered correctly by all students (zero discriminating power) but still are effective items. Since the items closely match all necessary learning outcomes, the results of the test simply tell us that here are outcomes that all students have mastered. Difficulty of an item in CRMM is determined by the learning task that the item is designed to measure. If the task is easy, the item should be easy. If the task is difficult, the item should be difficult. No attempt should be made to eliminate easy items or to alter item difficulty simply to obtain a spread of test scores (Popham, 1978).

The S-P chart analysis is a non-parametric method. It is good for analysis of the class performance data and preferable to the traditional method which relies on the normal distribution of measurement (Sato, 1980). It is an analytical method in which human vision is effectively useful in analyzing individual student's item response patterns in classroom tests. It diagnoses student's performance, instruction, and quality of a test. One of its characteristics is that item-responses can be presented in a graphical mode. An anomalous item pattern may reflect an ambiguity or other deficiency in an item. An anomalous student pattern may indicate gaps in knowledge or erratic study. These anomalous patterns are quantified by the caution index CP_j for the j-th item and the caution index CS_i for the i-th student. The heterogeneity of the S-curve and the P-curve is expressed by D^*, the disparity coefficient. Passing rate P_j of an item is the number of students who answered right divided by the total number of students. Passing rate of an item which increases after instruction shows sensitivity of learning gain.

The purpose of this work is to show how to construct and validate an instrument to assess sublimation technique on the basis of specific assessment criteria. The instrument can assess how accurately the students perform, give correct grades, and diagnose what has been learnt and what has not, so as to provide remedial instruction when needed. This instrument also assesses students' levels of attaining sublimation technique.

METHODS

I. Construction of the instrument

Sublimation was treated as a process concept in this study (Herron, et al., 1977). The content analysis [the structrual task analysis (Gagné, 1987), the process concept analysis (Herron, et al., 1977), the hierarchical skill mapping (Novak and Gowin, 1984)], the evaluation analysis, and the two-dimensional table of specification were used in the analysis of the sublimation technique. These analyses established criteria which led to construction of an evaluation tool to assess the levels of the technique attained, namely, concrete, identity, classificatory, and formal levels derived from the CLD (Cognitive Learning and Development) theory (Klausmeier and Sipple, 1980) and the step-by-step learning model (Dumas-Carré and Larcher, 1987). The specific assessment criteria obtained in the process concept analysis point out what the learner must do, how he must do, and why he so performs, so that the tool can give correct grade and search out what mistakes the learner made and how he can improve his experimental skills. Sublimation is a multiple task which cannot be assessed by only one assignment. It must be assessed by a rational set of new assignments. CRMM of sublimation must assess not only the accuracy but also the appropriateness of the performance, i.e., whether the performance is the correct response to the particular inputs (cues). In the evaluation analysis, the main variables considered are the characteristics of the learners, the level at which the technique is to be attained, the assignment that will be used, and the amount of teacher guidance to be provided (Klausmeier and Sipple, 1980). These analyses established criteria which led to construction of an evaluation tool to assess the learning levels of sublimation attained. A set of successive new assignments conforming to the criteria for the four attainment levels were used to assess retention or transfer of the acquired technique, i.e., to assess repeating, reproducing, generalizing and discriminating abilities of the students. In the selection of the assignments many sublimation cases were collected, from which the rational set matching all the representative attributes was screened. This set of assignments was so constructed as to differentiate the students into the four learning levels.

Concrete level: Successfully performing sublimation of a typical case instructed. This case should be able to cultivate the student's ability in "recognition-repetition".

Identity level: Successfully performing sublimation of a new case comparable with that used in concrete level, with expected behavior. This case should be able to cultivate the student's ability in "identification-reproduction".

Classificatory level: Successfully performing sublimation of a rational set of ordinary cases which are selected with consideration on all representative attributes. The cases should be able to cultivate the student's ability in "generalization to all ordinary cases".

Formal level: Successfully performing sublimation of new cases with behaviors not encountered before. The cases should be able to cultivate the student's abilities in "generalization including unusual cases, knowing the limits of the technique, and discriminating nonexamples (some of which may be coordinate techniques)".

II. Validation of the instrument

The validation of test materials consisted of two important activities. The first was verification that the test "as written" assessed the critical elements of the technique being tested. The second was verification that the test was assessed fairly and consistently. The process of validation incorporated a review and an administration of the test materials on a trial basis. The method of validating CRMM for sublimation used positive correlations (1) between the results of pre-test/post-test 1 and post-test 1/post-test 2, (2) between the results of check-list and paper-pencil test, (3) between the task performance and test result, and (4) among the results of the rational set of assignments.

A process (Fig. 1) was devised to revise and validate CRMM for assessing performance of the learners in sublimation. In this process the Student-Problem (S-P) Charts (Sato, 1980, 1986, 1988) were used to revise and validate the evaluation tool. The S-P chart analysis is a nonparametric method to analyze the class performance data. It is preferable to the traditional test-score theory which relies on the normal distribution of measurement (Sato, 1980). One of its characteristics is that item-responses can be presented in a graphical mode. An anomalous item pattern may reflect an ambiguity or other deficiency in an item. An anomalous student pattern may indicate gaps in knowledge or erratic study. These anomalous patterns were quantified by the caution index (CP_j) for the j-th item and the caution index (CS_i) for the i-th student. The heterogeneity of the S-curve and the P-curve was expressed by D^*, the disparity coefficient. Passing rate (P_j) of the j-th item was the ratio between the number of students who answered right and the total number of students. Passing rate of an item which increases after instruction showed sensitivity of learning gain.

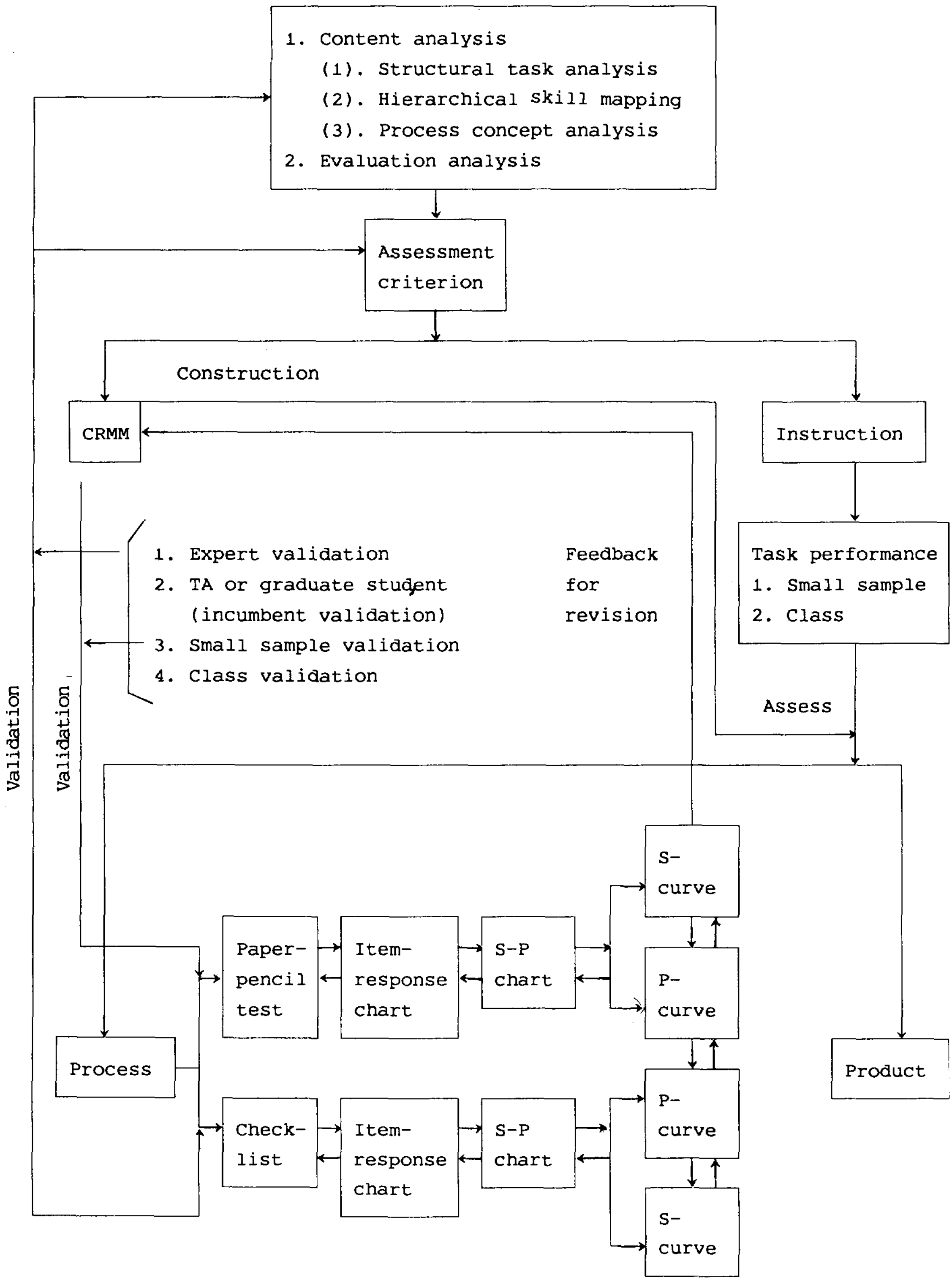

Fig. 1. Flow chart of validation process.

RESULTS AND DISCUSSION

I. Construction of the instrument

I-1. Structural task analysis

The steps in sublimation were identified as shown in Fig. 2.

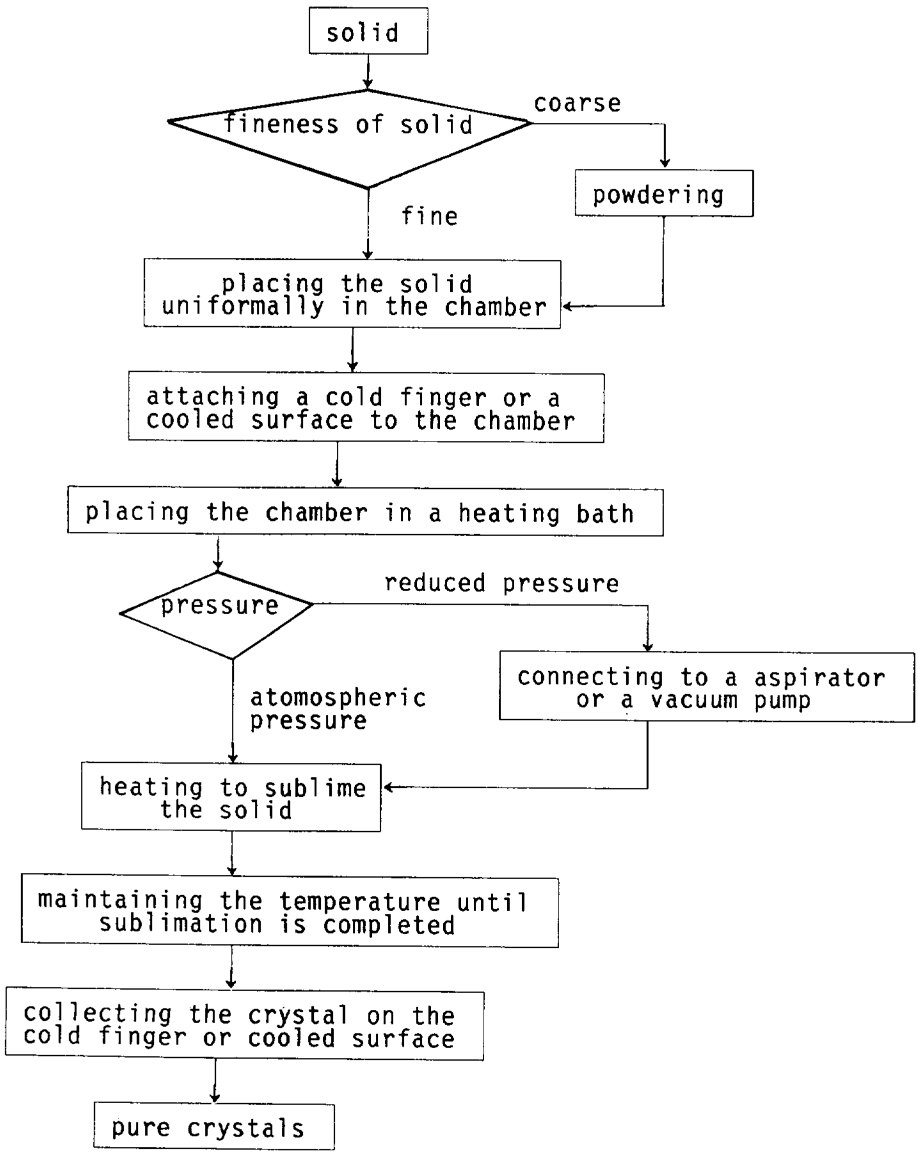

Fig. 2. Structural task analysis of sublimation.

I-2. Process concept analysis

Process concept analysis identified the following for the technique.

(1) Critical attributes

Placing the impure solid in the sublimation chamber and heating it to vaporize at a temperature higher than that of the surface on which the pure materials will be condensed but lower than the melting point of the solid.

(2) Variable attributes

a. Sublimation may be performed under atmospheric pressure or reduced pressure.

b. The substance to be sublimed may be a single solid or a mixture.

c. The coolant may be gas, liquid or solid.

(3) Superordinate, coordinate, and subordinate process concepts and their taxonomical inclusive-exclusive relationships are shown in Fig. 3.

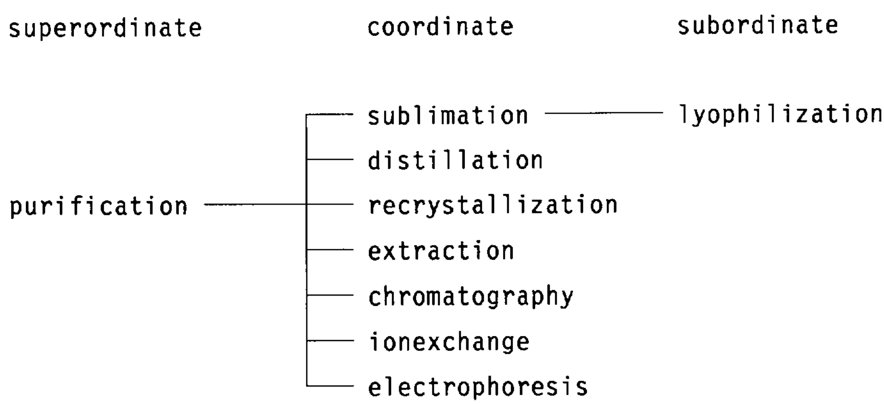

Fig. 3. Taxonomical analysis of sublimation.

I-3. Hierarchical skill mapping

Hierarchical skill mapping revealed skills prerequisite to sublimation and identified how sublimation was related to what the students already knew from their past experiences. The students were assessed for their ability in performing all the prerequisite skills. Those students lacking in any prerequisite skills were asked to make them up. The hierarchical skill mapping of sublimation is shown in Fig. 4.

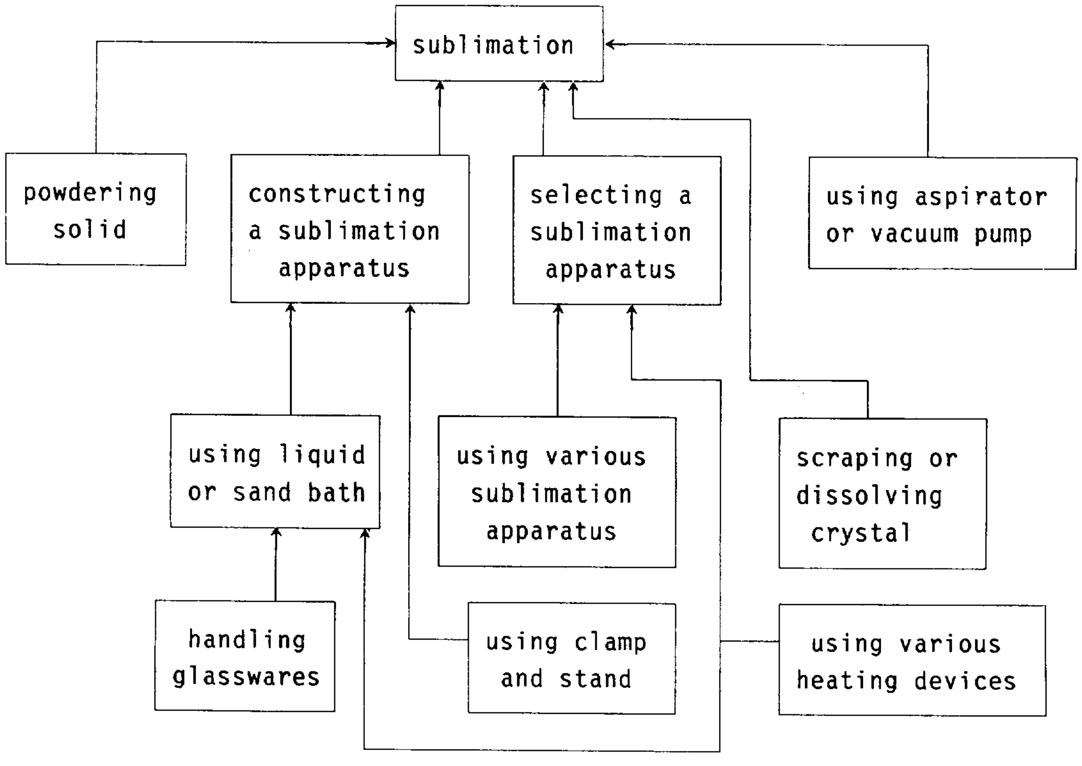

Fig. 4. Hierarchical skill mapping of sublimation technique.

I-4. Assignments for evaluation identified by the evaluation analysis

Concrete level: An impure solid such as p-dichlorobenzene for which the student was instructed.

Identity level: A solid comparable with that used in the concrete level, such as hexachloroethane, with expected behavior.

Classificatory level: A rational set of impure solids, such as anthraquinone, camphor, caffeine, salicyclic acid, 1-naphthol, and naphthalene, which sublime normally.

Formal level: Unusual cases. Outside of the scope of this work.

I-5. Two-dimensional table of specification

The two-dimensional table of specification with respect to the critical attribute of sublimation was constructed. One dimension of the table is for the skill category, and the other for performance category. With these specifications, performance criteria were written so as to match with the two categories.

I-6. Construction of check-list and paper-pencil test

The assessment tool in the forms of check-list and paper-pencil test was constructed from the two-dimensional tables of specification. The check-list contains the performances that an individual student must be observed by a teacher. The paper-pencil test takes the form of multiple-choice items; it is inexpensive, takes less time than the check-list, and is good for testing mental discrimination.

II. Validation of the instrument

II-1. Content validity

Content validity was determined by thorough inspection of the contents and the skill domains by a team of three experts. The expert judgment was based on whether all the behaviors listed in the two-dimensional Table of Specification were included in the test. Content validity must also test the appropriateness of the rational set selected. The content and the learning outcomes included in the test are based on the leading textbooks (e.g., Fieser and Williamson, 1987; etc.) and the recommendation of experts.

II-2. Item analysis

Learning achievement and its evaluation are affected by three factors: (1) student's previous experience, willingness to learn, and ability to learn; (2) the curriculum structure and the teaching method; and (3) the validity and the reliability of the evaluation tool. The third is influenced by the other two. Therefore, an effective validation procdure is needed to ensure the third factor in the presence of the influences from the other two. The validation procedure (Fig. 1) involved the item-response chart analysis and the S-P chart analysis combined with the check-list and the clinical interviews with freedbacks and revisions. Thus, a good evaluation tool led to high validity and reliability, which were confirmed by excellent agreement between the test result and the task performance as discussed in II-3-2-2 (below). Three teaching assistants served as incumbents.

II-2-1. Item validity

Item validity was observed from the downward shift of that item's portion of the P-curve in the S-P charts. Downward shifts represent sensitivity and uniformity of learning gains. Downward shifts were indeed observed in comparing the S-P charts for the results of the paper-pencil tests: (a) post-test 1 (Table 2) vs. pre-test (Table 1) and (b) post-test 2 (Table 3) vs. post-test 1. Item validity was further supported by four more evidences: (c) the check-lists and the clinical interviews; (d) P_j for each item increasing from pre-test to post-test 1 to post-test 2 as in P_j for item 13 increasing from 45% to 90% to 100%, showing sensitivity of learning gain; (e) CP_j of post-test 2 for sublimation being all less than 0.45, meaning all items are good (CP_j < 0.50) (Sato, 1980); and (f) in the plot of P_j vs. CP_j, all items being in Quadrant A, meaning good quality (Fig. 5) (Sato, 1980).

Table 1. S-P chart of sublimation - pre-test.

PROBLEM NUMBER D*=0.6329

Student No.	01	03	15	12	05	14	16	08	07	10	13	11	02	09	04	06	C.A.	%	CSi
11	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	13	81	0.09
19	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	13	81	0.70
6	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	12	75	0.03
3	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1	12	75	0.38
20	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	11	68	0.10
12	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	0	11	68	0.30
14	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	11	68	0.44
1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	11	68	0.51
18	1	1	1	1	1	1	0	1	1	0	0	0	1	0	1	0	10	62	0.24
21	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	1	10	62	0.33
8	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0	10	62	0.42
9	1	1	0	1	1	1	0	1	1	0	1	1	0	1	0	0	10	62	0.45
7	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	9	56	0.08
13	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0	9	56	0.25
2	1	1	1	1	1	0	1	0	1	0	0	0	0	1	1	0	9	56	0.37
4	1	1	1	1	0	0	0	0	1	1	0	1	1	1	0	0	9	56	0.65
15	1	1	1	0	0	1	1	0	0	0	1	1	0	1	1	0	9	56	0.74
16	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	8	50	0.05
10	1	1	1	1	1	0	0	1	0	0	1	0	0	0	1	0	8	50	0.36
17	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	7	43	0.02
5	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	7	43	0.30
22	1	1	0	1	0	0	0	0	1	0	0	1	0	1	0	0	6	37	0.62
A.C.	22	21	19	19	18	17	14	13	13	11	10	10	09	09	08	02			
P_j(%)	100	95	86	86	81	77	63	59	59	50	45	45	40	40	36	09			
CP_j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	.00	.32	.64	.96	.27	.54	.52	.50	.94	.30	.49	.67	.37	.87	.75	.61			

Table 2. S-P chart of sublimation - post-test 1.

PROBLEM NUMBER D*=0.4801

Student No.	01	05	07	12	03	08	13	14	15	16	04	10	11	02	09	06	C.A.	%	CSi
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	15	93	0.00
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	15	93	0.00
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	15	93	0.40
9	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	15	93	0.60
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	14	87	0.00
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	14	87	0.00
11	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	14	87	0.12
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	14	87	0.12
8	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	13	81	0.00
16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	13	81	0.00
15	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	13	81	0.20
18	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	13	81	0.20
5	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	13	81	0.40
19	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	13	81	0.40
21	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	13	81	0.55
13	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	12	75	0.00
2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	12	75	0.16
22	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	8	50	1.17
A.C.	22	22	22	22	21	21	21	21	21	21	20	17	15	15	13	09			
P_j(%)	100	100	100	100	95	95	95	95	95	95	90	77	68	68	59	40			
CP_j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.13	.55	.48	.67	.71	.35			

Table 3. S-P chart of sublimation - pre-test 2.

PROBLEM NUMBER D*=0.4136

Student No.	01	05	07	08	12	13	03	14	15	16	04	10	02	11	09	06	C.A.	%	CSi
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	100	0.00
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	15	93	0.00
17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	15	93	0.00
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	14	87	0.00
14	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	14	87	0.07
15	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	14	87	0.07
18	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	14	87	0.07
19	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	14	87	0.37
8	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	13	81	0.00
16	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	13	81	0.00
5	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	13	81	0.53
21	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	13	81	0.53
13	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	12	75	0.00
2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	12	75	0.24
22	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	10	62	0.86
A.C.	22	22	22	22	22	22	21	21	21	21	20	20	16	16	16	11			
P_j (%)	100	100	100	100	100	100	95	95	95	95	90	90	72	72	68	50			
CP_j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.15	.45	.39	.39	.43	.26			

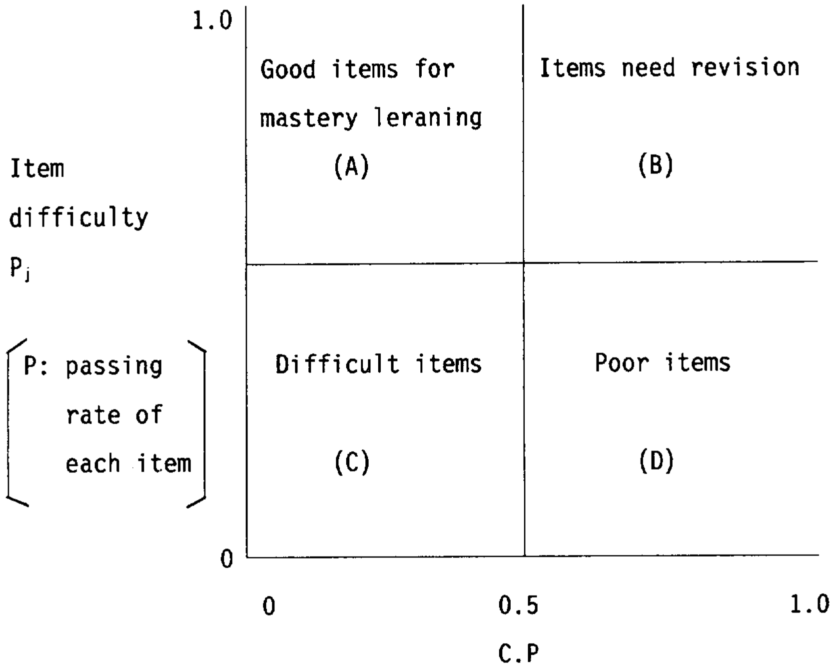

Fig. 5. P_j vs CP_j - Diagnosis of items for mastery learning.

II-2-2. Item reliability

Item reliability was observed by the rightward shifts of that portion of the S-curve (Tables 1, 2, 3), which showed consistency of the learning gain and the learning uniformity of each student on each test item when administering (a), (b), and (c) mentioned above (II-2-1), and (g) two equivalent items.

II-3. Criterion-related validity of the instrument

The instrument was validated on the basis of the assessment criteria which pointed out what, how, why, and when the learner must do specifically for sublimation. The validation procedure included the S-P chart analysis. The clinical interviews were additionally employed for those students having learning difficulties. The results were further confirmed by the positive correlations in (1)-(4) under METHODS.

II-3-1. Validation by the S-P chart analysis

The S-P charts were used to analyze students' response patterns in classroom tests. The S-P chart analysis is a formative/diagnostic evaluation method. Each item of the instrument for sublimation in the S-P chart was tied up to the criterion, i.e., the attainment target of the component skill. As the attainment approached the higher learning level, a higher degree of mastery of the component skill was observed. The P-curve as a whole was found to shift downward progressively from Tables 1 to 2 to 3, representing the sensitivity of learning gain. Validity of the instrument for sublimation was also observed in D*. D* decreased from 0.6329 (Table 1) to 0.4801 (Table 2) and finally to 0.4136 (Table 3), showing the sensitivity of learning gain. These facts indicate that the instrument is of good quality, because D* less than 0.5 is acceptable for multiple task (Sato, 1986).

II-3-2. Validation by expectancy table

The key elements in criterion-related validity are the degrees of agreements between the following two sets of measures, (1) the test scores and (2) the criteria to be predicted. Expectancy tables were used to show the relationship between a test score and an external criterion (Gronlund, 1982). In predicting success on the instrument, the categories were limited to mastery and nonmastery. The table shows, for each category on the predictor, how many students have demonstrated mastery or nonmastery on the external criterion. Degree of Agreement is 100 times (number of agreements in test score and external criterion)/(total number of students in the group).

II-3-2-1. Check-list and paper-pencil test

The performance on the check-list was highly related (95.5%) to the performance on the paper-pencil test for the post-test 2.

II-3-2-2. Process-product relationship

CRMM has high predictive validity, i.e., the performance on the test is highly related to the performance on the task (Campbell, 1986). The evaluation for the task, sublimation, was based on the product and the process. The product of the task performance is the yield and the purity of the crystals obtained. The process is assessed by the paper-pencil test. Those students (90.9%) with high scores in paper-pencil test indeed obtained the crystals in high yield and high purity. Thus the tool was confirmed to have high predictive validity.

II-4. Construct validity

Construct validity is related to the extent to which the proposed learning levels accounts for performance on the tasks of the rational set of activities having representative attributes.

Validation with respect to learning transfer after training was made by observing the individual performance profile (Table 4), which showed pass/fail in the rational set of activities. In Table 4, p-dichlorobenzene (the first column) was the first assignment (a typical case; the same task used in the instruction), which evaluated student's "recognition-repetition" ability. Students who failed in the first task were required to repeat the same task after a remedial instruction. Hexachloroethane (the second column) was the second assignment (a case comparable to the first), which evaluated student's "identification-reproduction" ability. Students who failed in the second task were given a third task selected from the rest of the rational set. Caffeine was the third assignment which evaluated student's "generalization to all ordinary cases" ability. The + signs show the successful performances. The last column records that 90.9% of the students reached the classificatory level; the rest remained in the identity level. These facts were further confirmed by the learning gain shown in the individual performance profiles constructed from the pre-test, the post-test 1, and the post-test 2 of the paper-pencil tests. Cosntruct validity was further confirmed by the paper-pencil test results, which showed the sensitivity and the uniformity of the learning gain of each student in going from freshman to sophomore years.

Table 4. Assignments and attainment level of individuals of sublimation.

Student no.	p-dichloro-benzene	hexachloro-ethane	caffeine	attainment level
1	+	+	+	classificatory
2	+	+	+	classificatory
3	+	+	+	classificatory
4	+	+	+	classificatory
5	+	+	+	classificatory
6	-	+	+	identity
7	+	+	+	classificatory
8	-	+	+	identity
9	+	+	+	classificatory
10	+	+	+	classificatory
11	+	+	+	classificatory
12	+	+	+	classificatory
13	+	+	+	classificatory
14	+	+	+	classificatory
15	+	+	+	classificatory
17	+	+	+	classificatory
18	+	+	+	classificatory
19	+	+	+	classificatory
20	+	+	+	classificatory
21	+	+	+	classificatory
22	+	+	+	classificatory

II-5. Reliability of the instrument for sublimation

The reliability of an evaluation tool is an indication of how a test consistently measures what it is supposed to measure. The degree of reliability of the tool is based on reliability of items within a test device. Item reliability of the tests used in this study was established in II-2-2.

II-5-1. Interrater reliability

The check-list contains the procedures that individual student must be observed for by the teacher. With four observers, the interrater reliability was ascertained by viewing videotapes and round table discussions.

II-5-2. Reliabiliity by the S-P chart analysis

The reliability was observed by the consistency of gain and the uniformity shown in the S-curve, which as a whole, shifted rightward progressively in going from Tables 1 to 2 to 3.

II-5-3. Comparing test performance on two equivalent halves of test items

The percentage of consistency (87.5%) was computed by the equation: 100 times (number of agreement in both halves)/(total number of students in the group).

CONCLUSION

Based on the findings and the discussions in the preceding sections, the following conclusions can be made.

1. Concept analysis was used in the analysis of sublimation to construct the evaluation tool which gave accurate grades and searched out the learners' mistakes to allow corrections by remedial instruction.

2. Content analysis and evaluation analysis were used to establish the criteria which led to the construction of the instrument to measure the learning levels of sublimation technique attained.

3. Content validity of the instrument was satisfactory for assessing whether students could perform sublimation of ordinary cases in college introductory organic chemistry laboratory course.

4. Item validity was clearly observed in the downward shifts of each item's portion of the

P-curve, showing the learning gain during a representative rational set of activities. CP_j values of all the items were less than 0.45. In the plots of P_j vs. CP_j, all the items were in Quadrant A. These facts indicated that all the items are of good quality. The item reliability was observed from the rightward shift of that portion of the S-curve showing consistency of the learning gain of individuals.

5. Criterion-related validity of the instrument was observed from the P-curve shifting as a whole downward in the three S-P charts. D* values were 0.4801 for post-test 1 and 0.4136 for post-test 2. The performance on the check-list was highly related (95.5%) to the performance on the paper-pencil test. Those students (90.9%) with high scores in the paper-pencil test indeed obtained the crystals in high yield and high purity.

6. Construct validity with respect to learning transfer was observed in the individual performance profile for (a) the task performances and (b) the test results in the rational set of activities.

7. The reliability of the instrument was observed in (a) the rightward shift of the S-curve as a whole and (b) 87.5% agreement in the two equivalent halves of items in the test, both (a) and (b) showing consistency of learning gain.

8. The procedure designed in this study involving the S-P chart analysis (a non-parametric method) was satisfactory in validating assessment instruments for classroom CRMM, as demonstrated by the application to sublimation, a chemistry manpulative technique.

ACKNOWLEDGEMENT

The authors would like to thank National Science Council, R.O.C., for financial support, and Dr. T. Sato of NEC, Tokyo, Japan, for providing information on the S-P chart analysis.

REFERENCE

Campbell, C.P. (1986). *Criterion referenced performance tests for measuring task proficiency.* Technological and Adult Education Department, The University of Tennessee.

Carver, R.P. (1974). Two dimensions of test, psychometric and edumetric. *American Psychologist* , 29(7), 512-518.

Dumas-Carre, A. & Larcher, C. (1987). The stepping of learning and evaluation. *International Journal of Science Education*, 9(1), 93-104.

Fieser, L.F. & Williamson, K.L. (1987). *Organic experiments.* 6th ed. D.C. Lexinton: Health and

Company.

Gagné, R.M. (1987). *Instructional technology: foundations.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Gronlund, N.E. (1982). *Constructing achievement tests.* 3rd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Herron, J.D., Cantu, L.L., Ward, R., and Srinivasan, V. (1977). Problems associated with concept analysis. *Science Education,* 61(2), 185-199.

Klausmeier, H.J. & Sipple, T.S. (1980). *Learning and teaching concepts. A strategy for testing application of theory.* New York: Academic Press.

Novak, J.D. & Gowin, D.B. (1984). *Learning how to learn.* New York, NY: Cambridge University Press.

Popham, W.J. (1978). *Criterion-referenced measurement.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Sato, T. (1980). The S-P chart and the caution index. *NEC educational information bulletin 80-1. C & C system research laboratories.* Japan: Nippon Electric Co., Ltd.

Sato, T. (1986). *Construction and interpretation of S-P chart.* 17th Printing. Tokyo, Japan: Mei-Ji Books Publisher (In Japanese).

Sato, T. (1988). *The S-P chart analysis.* Japan: NEC Technical College, NEC Corporation.

CONSTRUCTION AND VALIDATION OF INSTRUMENT TO ASSESS CHEMISTRY MANIPULATIVE TECHNIQUE - SUBLIMATION

化學實驗技能評量工具的設計與效化 —「昇華」技能

王澄霞　洪志明

化學系所

摘　　要

本文報告化學實驗技能 —「昇華」的標準參照評量工具之設計與效化過程，結果如下：(1)使用概念分析法分析「昇華」實驗技能，據以建立評量的基準，設計良好的評量工具，正確地評量出學生的成績，並找出學生犯錯之處，施以補救教學，而改進其實驗技能。(2)使用內容分析法及評量分析法，據以設計標準參照精通評量工具，來評量學生的學習達成階層。(3)評量工具的信度與效度適用於大學初級有機化學實驗課程之化學實驗技能的評量。(4)本研究所設計的含S-P表分析法（一種無母數法）之效化過程，能適用於效化班級標準參照精通評量工具，此由「昇華」實驗技能評量工具的效化而得到驗證。