

ソフトウェア品質の定量的評価技術

砂塚 利彦, 山岸 典子, 東 基衛
日本電気株式会社

1. はじめに

ソフトウェア品質の定量化の動きは数年前から米国を中心に起こり、1976年にBoehmが提唱した品質尺度が発端となり、ソフトウェアの総合品質に対するとらえ方の方向性が示された。その後も品質尺度についての研究開発が進み、新たな品質尺度の提案、品質保証体制の中への組み込みが始まり、多くの尺度・技法が混在しているのが現状である。また、品質尺度の側面である「信頼性」にスポットをあて、さらに掘り下げた研究が行われたり、ソフトウェア品質にかなりの影響を与える「複雑さ」の計測方法の研究なども進んでいる。

Boehmは60をこえる品質尺度を提唱したが、1977年にはWaltersとMcCallがRAD C (Rome Air Development Center) の報告書の中で品質のFactorを11項目に絞り込んだ。

1980年に発表されたMurine (METRIQS社) のSQM (Software Quality Metrics) は、それまでの数年間陸軍関係の企業で既に実施され、12項目のFactorも定量的に計測可能なレベルにまでブレイクダウンされた実用に向けた技術であった。

2. SQMの概要

SQMとは質のよいソフトウェアを安い費用で作ることを目的として、計測可能なFactor、CriteriaおよびMetricから構成されているソフトウェア品質の定量的評価技法である。図1に品質尺度の構造を示す。

SQMの品質尺度は、結果の品質を評価するだけでなく、品質を作り込むプロセスで計測管理できる演繹的尺度である。そのために段階的基準を設定している。

SQMの品質計測のアプローチは以下の通りである。

- ① 開発対象ソフトウェアとユーザの要求や開発者の方針などにより、要求される品質 (Factorという) を確認し、選定する。
- ② 各Factorを属性 (Criteriaという) の集合によってさらに定義する。
- ③ 各Criteriaを個々の計測項目 (Metricという) によって定量化する。

従って、各Metricを計測してCriteriaのスコアを算出し、それらを総合することによって保証するFactorを評価するというアプローチをとる。

Factorは12、Criteriaは23の中から選定し、Metricは必要なだけ設定する。計測は各Metricにつき該当する箇所をすべて抜き出し、その各々について品質要求を満足すれば「1」、満足しなければ「0」と評価する。

この品質計測を公式レビューの一環として各工程で実施して、評価が「0」の箇所に対してはProblem Reportを発行し、修正処置をとる。これによって次工程には悪さを送り込まないようにし、最終製品は品質の保証されたものができる。

このSQM技法は、主に軍関係のソフトウェアを開発している企業 (例えば、General Dynamics社) で採用され、充分に実績をあげている。

3. SQMの試行・評価

3.1 SQMの試行

SQMを実際プロジェクトに適用して方法論を評価するためのデータを収集した。ここでは3つのプロジェクトについて適用内容とその結果についての概要を示す。図2は各プロジェクトへの適用工程と評価対象の一覧である。

それぞれのプロジェクトの特性は以下の通りである。

プロジェクトAでは、ソフトウェアの開発と保守の支援ソフトを開発している。SQMを定義工程に適用して「使い易さ」を重点的に評価した。

Bでは原価管理システムを開発しており、設計・製造工程に適用してプロセスの品質を評価した。

Cでは交換機のOSの一部を開発している。ここには外注ソフトに対して受入検査の一環としてSQMを適用した。

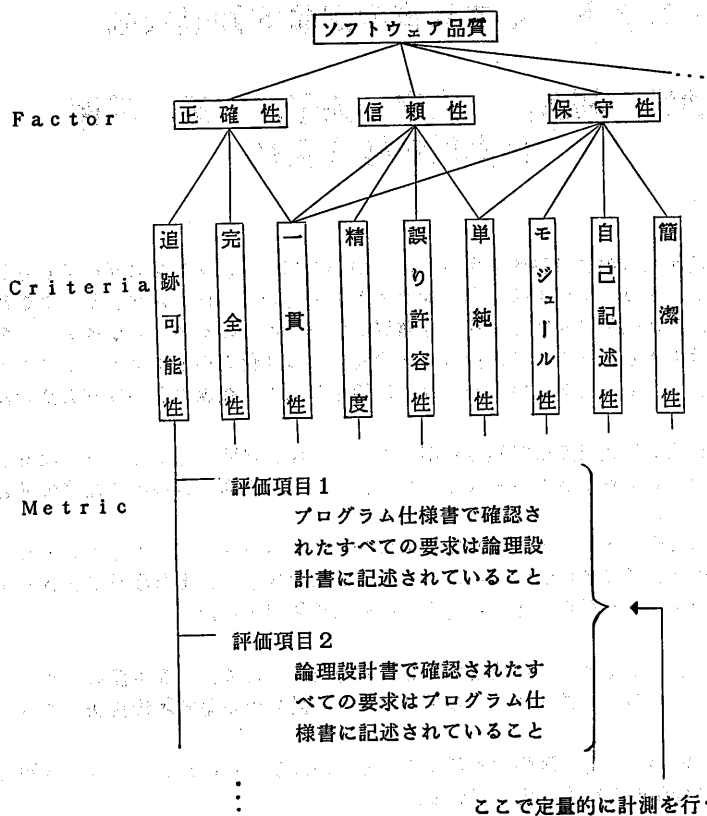


図1. SQMの品質尺度の構造

適用工程	分析定義	設計	製造	検査
プロジェクト			(C)	
A	↔			
B		HIPO	COBOL	
C		(SPチャート)	(PL/I)	↔

図2. SQMを適用した工程と評価対象

大項目 2	中項目 6	小項目 28
システム機能	機能自体の妥当性	- 効果 - 利用頻度 - ユーザ目標との適合性 ...
	他機能との連続性	- 作業の流れとの対応性 - データの連続性 - 相乗効果の程度 ...
	コストパフォーマンス	...
	システム目標	...
人間工学	人間特性	- 覚え易い - 親しみ易い - 親切さ - 誤り許容性 - 即時性 - 思考過程との連続性 ...
	インターフェース媒体	...

図3. プロジェクトAでの品質尺度

詳細を次に示す。

プロジェクトA

- ① 評価対象 : 要求仕様書
- ② 評価体制 : プロジェクト内にメンバー4人からなる評価チームを設置した。仕様書をその作成者以外のメンバーが客観的な立場から評価する。
- ③ 品質尺度 : 主な目標は「使い易さ」の実現であり、図3のような尺度を設定した。最下位のレベルでは28の評価項目がある。
- ④ 評価工数 : チームでの評価の前に、各メンバーは個人評価をしておく。個人評価には約1、2時間かかり、チーム評価には3、4時間かかった。
- ⑤ 実施効果 :
 - ・ 仕様書の枚数が半分に減り、読み易く容易に理解できるようになった。
 - ・ ユーザの要求に基づいた必要な機能だけを記述するようになった。
 - ・ 要求定義工程以降の開発工数が従来の3分の1に減った。

プロジェクトB

- ① 評価対象 : HIPOで記述された詳細仕様書およびCOBOLでコーディングされたソースリスト。4本のプログラムに適用した。
- ② 評価体制 : 評価グループはプロジェクト・リーダとプロジェクト外の第三者2名の計3名からなる。評価結果はリーダからプロジェクト・メンバーに報告される。
- ③ 品質尺度 : 各機能を完全に実現し、しかも信頼度をもたせるため、「正確性」「信頼性」を選択した。5つのCriteriaを選び(図4参照)、Metricは設計工程で15項目、製造工程で22項目設定した。
- ④ 評価工数 : 評価者が投入した工数は約3時間/0.5KLであり、全開発工数の15%であった。
- ⑤ 実施効果 :
 - ・ 検査工数がSQM未適用プログラムと比較して半減した(図5参照)。
 - ・ SQMを適用したプログラムからは、リリース後3ヶ月間にエラーは発見されなかった。もし未適用部分にも適用していたら、リリース後に発見されたエラーはあるMetricによって事前に検出できていたということが確認された。
 - ・ 設計工程での問題点は製造工程では改善された。従って総合品質スコアも84%(設計)から98%(製造)へと向上した。

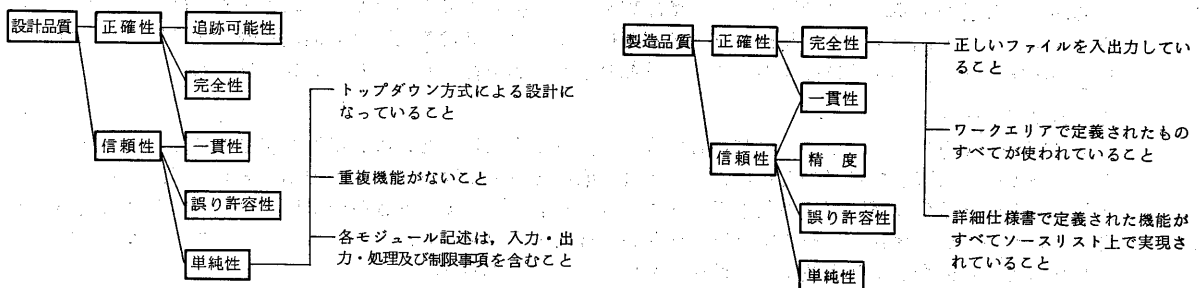


図4. プロジェクトBでの品質尺度

プロジェクトC

- ① 評価対象 : 外注ソフトに適用した。NEC側で「プログラム仕様書」を作成し、外注先はそれに基づいて「設計書」「プログラム」「ジョブ説明書」を作成した。受入検査としては従来「プログラム仕様書」および「ジョブ説明書」を基に「プログラム」のマシンテストを行っていたが、今回それに加えて「プログラム」の中味についてもSQMを用いて品質の評価を行った。
- 評価対象はPL/Iでコーディングされた「プログラム」である。
- ② 評価体制 : 評価者は発注元、外注先以外の第三者2名で、評価結果は発注元と一度検討した上で、発注元から外注先へ問題箇所をフィードバックした。
- ③ 品質尺度 : 受入検査の一環として「正確性」「信頼性」の評価を行った。Metricは12項目設定した。

品	リリース 3ヶ月後のバグ数
質	・ S Q M適用 0 件/K L
	・ 未適用 0.104件/K L
工	検査工数
数	49% 削除

図5. S Q Mの実施効果

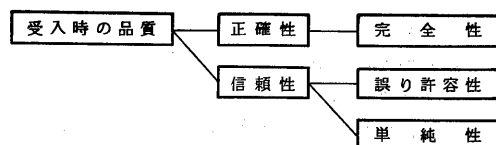


図6. プロジェクトCでの品質尺度

表1. 同一の計測対象に対する別々の計測者のスコア

	経験15年	2年目	新人
スコア	. 915	. 957	. 932
計測項目数	201	139	132
問題発見数	17	6	9
計測時間(分)	150	180	185
時間/計測項目数	. 746	1.295	1.402

表2. 検査結果とS Q Mスコアとの比較

比較項目 プログラム名	検査結果 (エラー数) 件 / K L	S Q Mスコア	スコアの内訳	
			正確性	信頼性
Prog-1	13. 9	. 944	. 905	. 986
Prog-2	21. 9	. 921	. 943	. 904
Prog-3	5. 3	. 974	. 963	. 988

④ 評価工数 : 工数はプログラム3本の平均で3. 4時間/0. 5 K Lであった。

⑤ 実施効果 :

- ・ 外注先のソフトのセールスポイント、ウィークポイントが明確になった。
- ・ 品質面において外注管理がきめ細かに行えるようになった。

3. 2 S Q M技法の評価

技法自体を評価するデータとして次の2種類を収集した。

① 同一の評価対象に対する別々の評価者のスコア(表1)

② 各評価対象の検査時のエラー数(表2)

それぞれデータの例を示す。

①のデータを分析することによってわかることは、評価者間のスコアの差が最大と最小とを比べても高々5%以内であるということである。また、評価者には新人からベテランまでいるので、誰にでも可能な客観的な評価であるといえる。

②からはS Q Mでのスコアの高いソフトは品質が良く、従って検査時に検出されるエラーは少なくなっている。逆にスコアの低いソフトは検出エラー数が多い。このように検査工程での品質評価とS Q Mでのスコアには高い相関があり、本技法は妥当性が高く、信頼できる評価を行っていることが裏づけられる。

4. S Q Mの改善

S Q Mを実際のプロジェクトに試行してみて、品質の定量化のための実用的な技法であることが確認できた。一方で、米国との背景の違いにより、評価対象はS Q Mの適用に向く大規模システムばかりではなかったり、品質保証はQ A部門だけの業務ではなく、全員参加のS W Q C活動の中で行うなど当社向けに改善すべき点もいくつか発見された。そこで当社ではS Q M技法の改善を行った。

当社の実情に適合する評価技術への要求条件は以下の通りである。

- ① ユーザ要求、ソフト種別に応じた尺度
- ② 定量的かつ客観的な品質指標
- ③ ライフサイクルの全工程で実施
- ④ 問題箇所の早期発見
- ⑤ 信頼度の高い評価(即ち、S Q Mのスコアと一般的に使われているエラーデータとの相関が高いこと)
- ⑥ 合理的な工数の範囲内で実施可能
- ⑦ 誰にでも可能な評価

これらを考慮した上で、試行結果から明らかになった問題点を整理すると次に3つに集約される。

① Factorの数(12)が多い。また類似のCriteriaで定義されるFactorがある。

人間が記憶できる項目の数には限界があり、10個以上はまず難しい。従って、Factorを選定する場合にはよく覚えている重要なものから常に選ぶようになりがちである。

また、いくつかの共通のCriteriaをもつFactorがある。例えば、Factorの「保守性」「テスト容易性」などは、前者は「一貫性」「単純性」「モジュール性」「自己記述性」「簡潔性」のCri

teriaで定義され、後者は「単純性」「モジュール性」「自己記述性」「計測性」で定義されている。実に3つのCriteriaが重複しているため、この2つのFactorを合わせて1つにして再定義することも可能である。

② ソフトウェアの種別によって要求される品質は異なる。従って各Factorの重要度には差がある。

例えば、不特定多数の人に使われるソフトであれば「使い易さ」と「正確性」を満足させる必要がある。金銭や機密情報を取り扱うシステムでは「セキュリティ」と「信頼性」に特に注意を払わなければならない。パッケージソフトであれば「柔軟性」と「保守性」を充分に考慮すべきである。

③ すべての項目を詳細に計測すると工数がかかりすぎる。逆にすべての項目を計測しないと、視点の偏った評価になる。

12のFactorの中から評価対象ソフトに必要なものを選択して品質計測を行っているが、選択されなかったFactorの視点からは全く計測は行われない。

そこで、これらの問題を解決し、より社内での実用に向くように改善を行った。

① Factorの絞り込み

② Factorの重要度の順位づけ

③ 品質計測方法の多様化（4段階評価の採用）

これらの具体的な改善内容を順に示す。

Factorの絞り込み

もともと12あったFactorを8つに絞り込んだ。類似のCriteriaをもつFactorは極力1つにまとめた。また、名称のつけ方も意味を理解しやすい表現に直した。

Factor	→	品質要求尺度
Criteria	→	品質設計尺度
Metric	→	品質評価尺度

改善後の品質要求尺度は表3の通りである。また、品質要求尺度と品質設計尺度との関係は表4のようになる。

また、絞り込んだ8つの品質要求尺度については、SQMでは選択して計測するが、改善後の方法では原則としてすべてを計測することにした。これによる計測工数の増加を避けるために、後述する重要度づけと計測方法の多様化を行った。

Factor（品質要求尺度）の重要度の順位づけ

8つの品質要求尺度をその重要度によって3段階に分けた。

特に重要	→	A
重要	→	B
普通	→	C

これはソフトウェアはその種別やユーザの要求によって優先的に満足すべき品質に違いがあるためである。

例えば、OA機器用ソフトであれば、「使い易さ」「正確性」「接続性」（複数の機能の統合活用）などが重要度Aに相当し、銀行オンラインシステムでは「信頼性」「セキュリティ」「正確性」などが第一に要求される。

この重要度づけにあたっては多くのパラメータが考慮される。ソフト種別、ユーザ要求、開発方針、規模、使用言語、ユーザ特性、難易度、支援ツール……。そして重要度A、B、Cの違いは次の計測方法に反映されている。即ち、重要な品質はどの程度に計測して高い精度で品質保証を行うというのが基本姿勢である。

品質計測方法の多様化

SQMでの詳細な計測方法に加えて、簡便な計測方法も採用した。それは4段階評価による方法であり、各計測項目に対して4から1のいずれかで評価する。

4	→	非常に満足
3	→	満足
2	→	ほぼ満足
1	→	不満足

SQMでは各Metric（品質評価尺度）に対して計測の該当箇所を列挙し、そのひとつひとつについて「満足」であれば1、「不満足」であれば0と評価して満足の割合でスコアを求めている。ここまで詳細に計測しなくともよい項目は4段階で評価を行う。

品質要求尺度の重要度と計測方法の関係について表5のように一応の目安を設けた。従来のSQMでは、このうち重要度がAの場合に相当する。B、Cの場合には計測していなかった。

以上、3つの改善を行うことにより、評価する品質の観点にもれがなくなり、また重要度に応じた計測方法であるため、計測工数もあまり増加せずにより適切な計測結果が得られるようになった。

表3. 品質要求尺度と品質設計尺度との関係

品質要求尺度 品質設計尺度	正確性	信頼性	保守性	柔軟性	使い易さ	効率	セキュリティ	接続性
追跡可能性	○							
完全性	○							
一貫性	○							
単純性		○	○					
精度		○						
誤り許容性		○						
モジュール性			○	○				○
自己記述性			○	○				
簡潔性			○					
計測性			○					
汎用性				○				
拡張性				○				
学習性					○			
対話性					○			
操作性					○			
マシン独立性				○				
ソフトウェア独立性				○				
実行効率						○		
記憶効率						○		
アクセス管理							○	
アクセス監査							○	
データ共通性								○
通信手順								○
インターフェース共通性								○

表4. 品質尺度の絞り込み

品質尺度(改善前)	品質尺度(改善後)
正確性	正確性
信頼性	信頼性
保守性	保守性
テスト容易性	→ 保守性
柔軟性	柔軟性
移植性	→ 柔軟性
再利用性	→ 柔軟性
使用性	使い易さ
効率	効率
保全性	セキュリティ
システム間動作性	→ 接続性
システム内動作性	→ 接続性

表5. 品質要求尺度の重要度と計測方法の関係

重要度	特に重要	重要	普通
計測方法	A	B	C
詳細計測	◎ 標準的	○ 評価工数が 充分にある	○
4段階評価 品質評価 尺度に 対して	× 望ましく ない	◎	○
品質設計 尺度に 対して	×	×	◎

5. 結 論

SQMは大規模プロジェクトに適用すれば非常に効果的である。そして次のような効果がある。

- ① 品質を定量化するため、具体的に品質目標が設定できる。
- ② 計測者間のスコアの差が小さいため、客観的評価ができる。
- ③ 計測結果をグラフ表示するため「目に見える管理」が行える。

さらに、試行結果をもとに問題点を整理し、当社の実情に適合するように改善して普及を進めた結果、以下の効果も得られた。

- ④ 多角的視点から評価するため、品質の共通尺度となる。
- ⑤ 簡便計測法(4段階評価)により中小規模のソフトにも合理的なコストで評価が可能となる。

その結果、プロセスの品質は向上し、プロダクトの品質は保証される。品質がよくなれば検査や保守の工数も大幅に削減でき、その分利益に反映される。

<参考文献>

1. Boehm, B.W. et al.(1976): "Quantitative Evaluation of Software Quality", 2nd ICSE, Vol.2, pp.592-605
2. Mizuno, Y.(1982): "Software Quality Improvement", 6th compsoc82
3. Murine, G.E.(1980): "Applying Software Quality Metrics in the Requirements Analysis Phase of a Distributive System", Proceedings from Minnow Brook Conference
4. Murine, G.E. et al.(1983): "Applying Software Quality Metrics", ASQC Quality Congress Transactions
5. Murine, G.E. et al.(1984): "Measuring Software Product Quality", Quality Progress, Vol.17, No.5, pp.16-20
6. 東 他(1984): 「ソフトウェア品質の定量的評価技法(Ⅰ) —— 要求条件と運用 ——」, 情報処理学会第29回全国大会講演論文集, p.681-682
7. 砂塚 他(1984): 「ソフトウェア品質の定量的評価技法(Ⅱ) —— 尺度と測定方法 ——」, 情報処理学会第29回全国大会講演論文集, p.683-684
8. 山岸 他(1984): 「ソフトウェア品質の定量的評価技法(Ⅲ) —— 事例と評価 ——」, 情報処理学会第29回全国大会講演論文集, p.685-686